**The Dissertation Committee for Peirong Lin certifies that this is the approved version of the following dissertation:**

# Towards Actionable Climate and Flood Prediction: Understanding and Advancing Land Surface Modeling with Enriched Geospatial Information

**Committee:**

Zong-Liang Yang, Supervisor

M. Bayani Cardenas

Robert E. Dickinson

Randal D. Koster

Jiangfeng Wei

# Towards Actionable Climate and Flood Prediction: Understanding and Advancing Land Surface Modeling with Enriched Geospatial Information

**by**

**Peirong Lin**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**May, 2018**

# Dedication

This dissertation is dedicated to my parents and sister

who gave me a family of love to start with

and to my husband Kai Zhang

who makes me happy and grows with me

## Acknowledgements

Upon completion of this dissertation, I would like to take this opportunity to express my sincere thanks to all the mentors and friends who accompanied me during this journey.

First and foremost, I would like to thank my supervisor Dr. Zong-Liang Yang for all his generous supports that make this happen. He is so selfless and is always willing to spend hours to share his broad knowledge in earth sciences, new thoughts on research, and career advices with his students. I benefited a lot from many in-depth discussions with him on a broad range of interdisciplinary topics, which helped me to think outside the box as well as to have critical thinking. I am grateful he told me that "the sky is unlimited" when I was initially worried about my limited understanding in atmospheric science and climate models when first entering UT. He encouraged me to keep learning all courses and tools needed to fulfill my research interests, while offering advices wherever needed. He is a very patient, understanding, supportive, and knowledgeable advisor. My Ph.D. study has been very colorful because of his kind offers to all possible opportunities that may make my career bright.

I would like to thank Dr. M. Bayani Cardenas for the insights he provided and his motivating teaching style that implants curiosity. His class was always a joy to me as he usually explains the fundamental concepts and quantitative parts of geoscience in a very clear way. Each time I talked to him, he would quickly grasp what I was doing and drop beneficial comments to help me rethink my research. I am grateful that he urged me to "better understand this world", which I strove hard to achieve and will continue to think it through in my academic career.

I would like to thank Dr. Robert E. Dickinson for all the advices he provided to my research work and the tips to improve my writing. He gave me lots of encouragement when I was a first year Ph.D. student in his class on *Climate Change: Current Literature Review*. Throughout these

years, he offered me continuous supports and discussions to my research whenever I was in need. Kai and I have been very grateful to the generous advices that he and Dr. Rong Fu gave us both in career and in life.

I would like to thank Dr. Randal D. Koster for kindly agreeing to serve on my dissertation committee and taking the time to discuss my research projects with me several times over the phone, even though he is at NASA. He has been extremely responsive in offering me critical and insightful comments, which keeps urging me to think about how to conduct better research. Before I first got the chance to talk to Dr. Koster in AGU 2015, I had been following his research papers to learn his great thoughts in my area of study. I am very grateful that he is willing to take time to interact with young researchers like me, and offering his suggestions.

I would like to thank Dr. Jiangfeng Wei for sharing his research experiences with me and guiding me through several difficulties. I benefited a lot from his class on *Coupled Earth System Modeling*, in which he introduced the science of coupled climate models, gradually guided us to read literature, find caveats, and then come up with a doable class project. The research initiated in this class eventually flourished and became Chapter 2 of this dissertation.

In addition to my dissertation committee, I would like to express my special thanks to Dr. David R. Maidment for the mentorship he provided to me. The work presented in Chapters 4 to 6 has benefitted a lot from my interactions with Dr. Maidment's research group. I took his GIS class in 2014, but it was since 2016 that I started to regularly attend his research group meetings and closely engaged in the discussions to learn his insights, passions, and great leadership in hydrology. He helped me realize that flood emergency response, the eventual piece of my research goals, is not as easy as we might thought, because huge gaps still remain to translate science and technologies to the public. He kindly offered chances for me to be the observer in the Austin Fire Department and the Texas Division of Emergency Management, which transformed the way I

think about how science can be better conveyed to address the specific needs in an actual flood emergency. He and Dr. Yang encouraged me to serve as the course coordinator of the 2016 Summer Institute of the National Flood Interoperability Experiment (NFIE) held at the National Water Center (NWC), which is an asset to me due to the new forms of education and research involved. I had a fun time discussing Chinese history with Dr. Maidment and helping him translate his PowerPoint slides into Chinese characters to better convey his messages to researchers in my home country. I also want to extend my thanks to Dr. David Arctur, Dr. Timothy Whiteaker, Dr. Paola Passalacqua, Harold R. Evans, and Xing Zheng (Dr. to be) for helpful discussions when I joined the research meetings at CWE.

For my research, I would also like to thank Dr. David J. Gochis and Dr. Wei Yu for hosting my visit to the National Centers for Atmospheric Research, and teaching me the skill in running WRF-Hydro. Dr. Adnan Rajib is thanked for all the helpful discussions we shared about hydrologic modeling, and the productive collaborations we had at NWC. During the past two years, I have also benefited greatly from Dr. Larry J. Hopper Jr., Jon Zeitler, and Mark Lenz from NWS, NOAA in a close research collaboration that we had. Thank you to them for being great collaborators and for helping me better understand the field of hydrometeorology.

All members of the Land Atmosphere Environment Dynamics (LEAD) group are also greatly appreciated. Dr. Hui Zheng and Dr. Armaghan Abed Elmdoust are both good friends and motivating colleagues to me; discussions with them have been always joyful and insightful. I also benefited a lot from discussions with other previous LEAD and UT members, including Dr. Cedric David, Dr. Hua Su, Dr. Mingjie Shi, Dr. Xitian Cai, Dr. Yongfei Zhang, Dr. Qinjian Jin, Dr. Bing Pu, Dr. Long Zhao, Dr. Yonghwan Kon, Dr. Sagar Parajuli, Dr. Ahmad Ali Tavakoly, Dr. Zhongfeng Xu, Dr. Mingxing Li, Dr. Marcelo A. Somos-Valenzuela, and Dr. Muhammad Shaikh. I would also like to thank the current LEAD members including Wen-Ying Wu, Lingcheng Li, Maryia Halubok, Jingjing Liang, and Seungwon Chung.

Special thanks go to Jingjing Liang for being a good roommate and cooking me good food when I was busy with dissertation.

Several persons at the Jackson School of Geosciences also helped me during these years. I am grateful that Philip A. Guerrero has been always responsive in solving our problems. I am grateful that Dr. Ginny Catania offered the class on *Preparation for Future Geoscience Faculty* and leads the efforts to initiate discussions among women scientists. I am grateful to Dr. Kerry Cook for being a great teacher to help me understand *Climate Dynamics*; I enjoyed a lot when being her Teaching Assistant in *Global Warmin*g. I am also grateful that Dr. Dan Breecker, Dr. Julia Clarke, and my advisor co-thought the very insightful class on *the Asian monsoon*. I have never sit in a class with three professors showing up at the same time, and freely allowing the discussions on any classic papers and unresolved issues for monsoon studies from interdisciplinary perspectives. This class was quite an eye-opening experience for me, and I am glad I took this class despite my busy schedule in my last semester at UT.

I thank Lu Pan for being my best friend and always the role model for me.

Lastly, I am indebted to my family who gives me endless love and supports.

**Towards Actionable Climate and Flood Prediction: Understanding and Advancing Land Surface Modeling with Enriched Geospatial Information**

Peirong Lin, Ph.D.

The University of Texas at Austin, 2018

Supervisor: Zong-Liang Yang

Land surface models (LSMs) are central to our understanding and prediction of the terrestrial hydrological cycle. This dissertation focuses on using enriched geospatial information from remote sensing (RS) and geographic information system (GIS) to advance the snow and river routing component of state-of-the-art LSMs, and assessing their roles in predicting temperature, precipitation, and streamflow.

In Chapters 2 and 3, the first systematic studies are conducted to quantify the role of land snow data assimilation (DA) in seasonal climate forecast. Using 7-yr DA products that assimilated the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) and the Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS), I find a local improvement of 5%–25% in the temperature forecast, where the delayed improvement at higher latitudes is explained by incoming solar radiation that is key to the snow–atmosphere coupling. Focusing on the Asia monsoon, I detect an improvement in the precipitation forecast, which is more robust over central north India with sensor-dependent behaviors in different seasons. The results clarify that to successfully translate DA to useful atmospheric prediction skill, the regional snow–atmosphere coupling, the DA uncertainties, and the monsoon sensitivity to thermal forcing over land need to be jointly considered. In Chapters 4 and 5, I introduce a vector-based river routing model to be coupled with traditional grid-based LSMs. By conducting comprehensive model evaluations in the Texas "Flash Flood Alley" in high-impact historical floods, I identify the model strengths and weaknesses in simulating flood discharges. The best modeling results are then

used to reveal the hydrometeorological factors responsible for a record-breaking local flood, which includes the rainfall location and basin physiographic features, the initial wetness in the deeper soil layer, and the flow velocity in the river network.

The assessed modeling advancements have actionable societal implications because they apply to the Community Land Model 4 (CLM4) and the Noah model with multi-parameterizations (Noah-MP), both LSMs are adopted by major operational forecasting centers. They may also inform future LSM developments that aim to unify the "top-down" atmospheric modeling and the "bottom-up" hydrological modeling approaches in a generic framework.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

*Land* is a key earth system component where human lives and complex biogeophysical and biogeochemical processes happen. Land surface modeling is central to our understanding and prediction of the coupled terrestrial hydrological cycle, which has direct societal implications. This dissertation aims to achieve a better fundamental understanding on the role of *land* in hydroclimate predictions. From the viewpoint of enriched *geospatial information*, it also summarizes several of my research endeavors to improve the prediction for temperature, precipitation, and streamflow, all of which necessitate new features to be introduced in traditional land surface modeling. In this chapter, I will first review the land surface model (LSM) evolution and the continuing need for geospatial information in characterizing land surface conditions while adapting LSMs for actionable uses, upon which this dissertation develops.

## 1.1. GEOSPATIAL INFORMATION IN LAND SURFACE MODELS

Land, atmosphere, ocean, and sea ice are the four major physical components of the Earth's climate system. Among them, land has the most pronounced spatial heterogeneity, making the adequate modeling of the energy and water flux partitioning at the land surface a non-trivial research task (e.g. Dickinson 1995; Giorgi and Avissar 1997).

The LSM development of the past 50 years is a gradual process to account for improved model physics as well as better representations of land surface heterogeneity (**Fig. 1.1**). The broad spectrum of today's sophisticated LSMs can be traced back to the 1$^{st}$ generation simple "bucket" model originated from the atmospheric modeling community (Manabe 1969) to provide an interactive lower boundary for the coupled ocean–atmosphere general circulation models (GCMs). To better represent the fluxes feeding back to the atmosphere, later LSM developments gave vegetation a more direct treatment due to its strong controls on these fluxes, which marked the 2$^{nd}$ and 3$^{rd}$ generation LSMs to connect with the ecosystem and carbon modeling (Deardorff 1978; Dickinson et al. 1993; Sellers et al. 1986). The LSM development philosophy coincided with that

of the hydrological modeling community (Eagleson 1986; Wood 1991) in addressing the need to represent subgrid variability for coarse resolution LSMs. This motivated studies to use natural topographic boundaries to discretize the land surfaces (e.g. Koster et al. 2000b) and to improve the parameterizations of runoff, snow, groundwater processes (Liang et al. 1994; Maxwell and Miller 2005; Niu and Yang 2007; Niu et al. 2007; Pitman 2003; Yang 2004). The most recent LSM advancements include more detailed ecological and biogeochemical processes (Bonan and Levis 2006; Cai et al. 2016; Sun et al. 2014). In the meantime, the plethora of LSMs has also prioritized model inter-comparison projects assessing different hydrologic variables (Henderson-Sellers et al. 1993; Scanlon et al. 2018; Xia et al. 2012) and model benchmarking (Best et al. 2015).

Among the numerous pathways for LSMs to move forward, this dissertation is particularly interested in the trend for LSMs to become less distinct from those of the hydrological origin, which still requires more LSM developments to advance our understanding, modeling, and predictive capability of the hydrological cycle over large spatial domains and at fine resolutions. More specifically, I focus on enriching the *geospatial information content* of the LSMs to improve the prediction skill for temperature, precipitation, and streamflow.

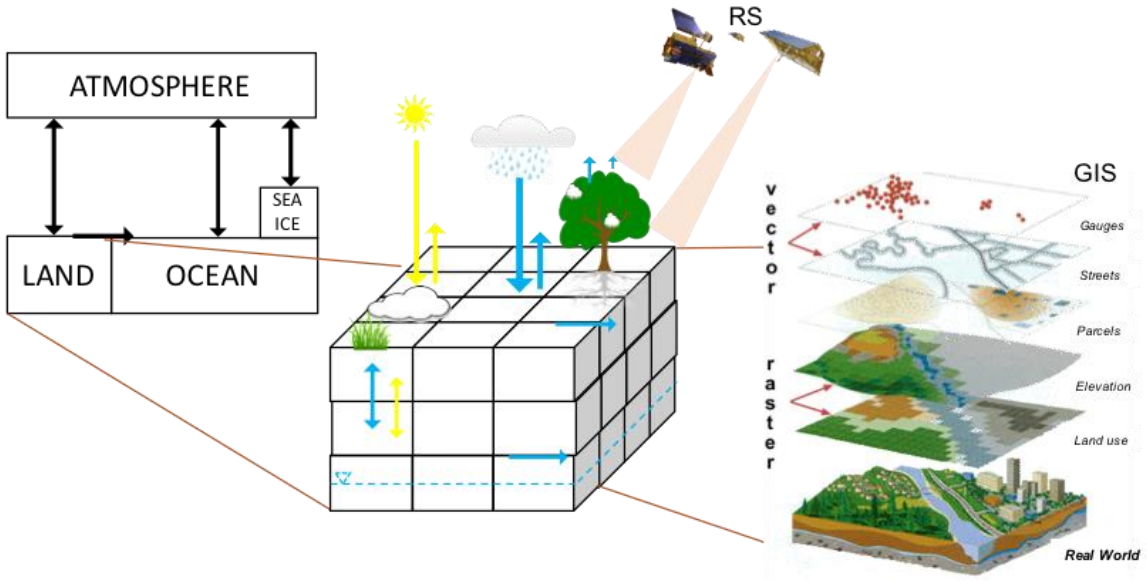


**Figure 1. 1.** Schematic diagram of the LSM development towards real-world spatial heterogeneity. Geospatial information content in LSMs is low to high from left to right. For simplicity, only selected LSM processes are shown in the middle. GIS layers are from: https://serc.carleton.edu/eyesinthesky2/week5/intro_gis.html.

Since the 1960s to 1970s, the emergent *geospatial information techniques* have transformed the way we perceive land surface heterogeneity. Digital mapping of global land surfaces by remote sensing (RS), global positioning system (GPS), and geographic information system (GIS) has greatly enriched the *geospatial information*, i.e., data and attributes about *place* and *location*, for use in modeling and understanding of the real-world spatial heterogeneity. Nowadays, Geographic Information Science (*GIScience*) not only refers to the *techniques* for geospatial data acquisition, storing, and visualization, but it also emphasizes the *science to unravel the geographical and dynamical patterns associated with "where"* (Goodchild 1992). While this is in line with the "hyperresolution" goal of the LSM development (Wood et al. 2011), numerous gaps remain for LSMs to ingest such geospatial information and to serve the goal of improving the understanding and prediction of the coupled hydrologic cycle. For example, current LSMs only utilize geospatial data such as global land cover, vegetation type, greenness, and soil texture as the *static* data inputs, while a wide variety of established geospatial data fabrics depicting fine-scale hydrological and societal features such as river channels, catchments, gauges, reservoirs, building address points, streets, and urban drainage systems have been limited in use. In addition, *dynamical* strategies to combine satellite remote sensing (discontinuous in space and time) with LSMs (continuous in space and time), such as data assimilation, have only emerged as an area of active research in recent decades for methodological development.

In this dissertation, I specifically focus on two overarching questions. *First*, to what extent can satellite remote sensing improve our characterization of global land surface conditions and what roles does it play in atmospheric predictions through land–atmosphere interactions? *Second*, how can we take advantage of the hydrologically-relevant geospatial data such as channel network, gauge, and catchment to improve streamflow prediction and aid decision support?

## 1.2. MOTIVATING RESEARCH QUESTIONS

These overarching questions are addressed with two important applications of the LSMs. The first piece (Chapters 2 and 3) is on *seasonal climate prediction* targeting at a lead time of one to several months, in which the memory of land snow plays a key role. The second piece (Chapters 4 and 5) is on *large-scale hydrologic prediction* for streamflow.

### 1.2.1. Simulation, prediction, predictability

It is important to caution the use of *simulation*, *prediction*, and *predictability* before this dissertation develops further. Although I aim to gain insights into a range of *prediction* problems for temperature (Chapter 2), precipitation (Chapter 3), and streamflow (Chapters 4&5), the experiments conducted are essentially model *simulations* that largely simplify the full range of complexity in making an actual prediction. For example, the seasonal "hindcast" experiments conducted in Chapters 2 and 3 have eliminated the potential impacts from sea surface temperatures (SSTs) and sea ice by prescribing them with the observations, while in an actual prediction setting, the SSTs and sea ice modules also need to be properly initialized and allowed to interact with the atmosphere, in order to mimic the natural climate system. The streamflow simulation experiments conducted in Chapters 4 and 5 can only inform actual real-time predictions if the rainfall inputs are obtained from weather forecasts, but I use radar-based rainfall to minimize the uncertainty from the weather forecast, in order to focus more on the hydrologic responses. Nevertheless, I will discuss the scientific basis and challenges associated with actual predictions in the following section, because these research efforts are all attempted to understand the *predictability* of different components of the hydrological cycle, where *intrinsic predictability* refers to the upper limit of the forecast skill, if exists, of a system, while *practical predictability* refers to the maximum possible forecast skill if the best-known procedures or models are utilized to make a prediction (Koster et al. 2004a; Lorenz 1996). These caveats necessitate careful interpretations of results, which will be briefly discussed in each chapter and summarized altogether in Chapter 6.

### 1.2.2. Seasonal climate prediction

Climate scientists have been actively seeking sources of atmospheric predictability (mainly temperature and precipitation) beyond the weather forecast time scale (Shukla 1998). Unlike weather forecast that has an upper limit of two weeks due to the atmospheric "chaos" (Lorenz 1996), seasonal climate forecast can potentially obtain predictable signals from the slowly varying ocean and land conditions through ocean–atmosphere and land–atmosphere interactions in the forecast period (Doblas-Reyes et al. 2013; Goddard et al. 2001). In particular, land contains a memory of weeks to months by storing past climate information in anomalous snow, soil moisture, vegetation, and groundwater conditions, which can influence the atmospheric prediction over seasons and areas beyond the controls of SSTs. Thus, land-derived seasonal climate predictability has received increased attentions in recent years (e.g. Douville 2010; Hirsch et al. 2014; Jeong et al. 2013; Koster and Walker 2015; Koster et al. 2004b; Koster et al. 2010b; Orsolini et al. 2013).

To obtain forecast skill from land, however, one of the most challenging tasks is to accurately characterize global land surface conditions at the time point of issuing a forecast, which is almost always impossible due to the uncertainties associated with any models or observational techniques to derive such an estimate. To this end, land data assimilation (DA) emerged as a promising statistical approach to combine the strengths of physically-based LSMs and observations (Lahoz et al. 2010). However, our scientific understanding on how much seasonal climate forecast skill can be derived from land DA is largely limited. Of particular interests to this dissertation is the land snow that can exert influences to the atmosphere through immediate albedo effect and delayed hydrological effect (Xu and Dirmeyer 2011; Yasunari et al. 1991). Despite numerous research efforts to obtain a robust estimation of global snow conditions by assimilating different satellite observations into LSMs (Kwon et al. 2017; Slater and Clark 2006; Su et al. 2010; Sun et al. 2004; Zhang and Yang 2016; Zhang et al. 2014), snow-oriented DA has not been utilized in seasonal climate prediction. Subsequently, the spatiotemporal pattern and the degree to which

snow DA can contribute to the atmospheric prediction is largely unexplored, which limits our understanding on the intrinsic and practical predictability of the seasonal climate forecast.

*This dissertation thus aims to address the following questions in Chapters 2 and 3:*

(1) Can snow DA, assimilating multi-sensor satellite observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Gravity Recovery and Climate Experiment (GRACE), improve the snow initialization and subsequently the seasonal temperature and precipitation forecast?

(2) If so, what are the spatiotemporal patterns and why?

(3) What are the factors potentially influencing the successful translation of the snow DA signal to improved seasonal climate prediction?

### 1.2.3. Large-scale flood prediction

Streamflow is one of the most important variables for freshwater availability, which has many practical implications. Streamflow patterns and dynamics over land can be simulated using an LSM coupled to a river routing model. However, numerous caveats exist to constrain our understanding and prediction of fine-scale streamflow dynamics over a large geographical domain, particularly for flood. First, previous model evaluations focusing on LSM-simulated streamflow were mainly performed at daily, monthly, and annual time scales, and only at limited major basin outlets over the globe (Cai et al. 2014; David et al. 2011; Li et al. 2013; Lin et al. 2018; Xia et al. 2012; Yang et al. 2011). These studies have satisfied the atmospheric modeling purposes from a "top-town" approach, however, such model–data comparisons are apparently not enough to fulfill the targeted "hyperresolution" modeling goal (Wood et al. 2011). Second, river routing models for use in earth system models were typically developed using a grid-based approach (e.g. Li et al. 2013; Lohmann et al. 1998), but the relatively coarse spatial resolution as limited by computational power is constraining the fine-scale depiction of hydrologic features. Subsequently, they have been limited in use to understand the problem of flood predictions. Third, most existing river routing

models were developed as stand-alone tools using LSM-simulated runoff as the forcing (David et al. 2011; Liu and Hodges 2014; Mizukami et al. 2016). However, to comprehensively understand the flood prediction, an integrated assessment on the uncertainty cascade from meteorology, to LSMs, and eventually to routing models is warranted, which has been lacking. These caveats have limited our modeling and forecasting capability for floods, as well as the local flood emergency response at the neighborhood scale.

*This dissertation thus also aims to address the following questions in Chapters 4 and 5:*

(1) Can GIS-based approaches for representing irregular modeling units be introduced to traditional grid-based LSMs to improve our modeling capability and facilitate large-scale fine-resolution hydrological predictions?

(2) What is the model simulation skill during high-impact historical flood events? Can we use the best observational constraints to identify the model strengths and weaknesses?

(3) What are the uncertainties associated with weather inputs, LSMs, and routing models? What factors are constraining our predictive skill for record-breaking local floods?

## 1.3. ORGANIZATION OF THE DISSERTATION

Chapter 1 reviews the LSM evolution and the continuous need for enriched geospatial information to obtain skillful hydroclimate prediction, followed by the overarching research questions of my dissertation work.

In Chapter 2, I investigate the role of snow DA in seasonal temperature forecast, by utilizing two suites of global snow DA products (2003 to 2009) in initializing the land component in over five hundred of ensemble-based seasonal "hindcast" experiments. By comparing the modeling results with global "observed" air temperature data from the Hadley Centre and the ERA-Interim reanalysis and contrasting them, I find an improvement of 5%–25% in temperature forecast locally over regions with prominent DA updates such as the Tibetan Plateau (TP), Siberia, and northern North America. Interestingly, we find that the improvement only appears later in

warmer months at higher latitudes while immediate forecast improvement could be observed at lower latitudes. A further regional analysis focusing on the TP and Siberia reveals that a lack of incoming solar radiation in cold months is the reason for the delayed response in Siberia.

Chapter 3 investigates the role of springtime snow DA in seasonal precipitation forecast. This chapter extends the model integrations conducted in the previous chapter to 6 months till the end of August each year, and the goal is to assess the Asian monsoon prediction skill derived from better snow initializations over the TP and Siberia. Focusing on the land-sea thermal contrast as the fundamental driver for monsoon, we find the most robust improvement signal is located in central North India, one of the strongest Asian monsoon subcomponents. We clarify the complexities in translating snow DA signals to improved rainfall forecast, which needs to consider the regional snow-atmosphere coupling strengths, snow DA uncertainties, and the monsoon's sensitivity to thermal effects of the Eurasian landmass.

In Chapter 4, I present the model coupling work that implements a vector-based routing model in the community WRF-Hydro framework. This chapter illustrates the importance of an alternative vector-based routing with traditional grid-based LSMs, followed by an application that simulates the inland riverine flood discharge during Hurricane Ike in 2008. The hybrid model's sensitivity to the land grid resolution, the coupling interface, and the flow travel time estimation is assessed to inform future research.

Chapter 5 further extends the work in Chapter 4 by augmenting the model to best resemble the United States National Water Model (NWM), which became operational in August 2016. By focusing on the Texas Hill Country, a well-known flooding "hotspot" of the continental United States (CONUS), I systematically evaluate the flood prediction skill of the model during two high-impact flood events with all best available observational constraints. In addition to model evaluation, the best-simulated modeling results are also used to gain hydrometeorological insights into a local record flood. The results point to three reasons for the relatively unexpected record-breaking flood at Blanco River at Wimberley in May 2015. These include the location of the intense rainfall combined with basin physiographic features, the subsurface soil saturation degree,

and the spatial combination of the flow velocity field in the river network, all of which making the successful prediction of such a record flood challenging.

Chapter 6 presents a regional flood inundation mapping application extended from my research work for use in real-world flood emergency response. Major conclusions, limitations, and future research work are then summarized.

The bibliography for individual chapters will be listed altogether in the end.

# Chapter 2: Snow data assimilation-constrained land initialization improves seasonal temperature prediction[1]

## 2.1. ABSTRACT

We present the first systematic study to quantify the impact of land initialization on seasonal temperature prediction in the Northern Hemisphere, emphasizing the role of land snow data assimilation (DA). Three suites of ensemble seasonal integrations are conducted for coupled land–atmosphere runs. The land component is initialized using datasets from (1) no DA, (2) assimilating the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF), and (3) assimilating both MODIS SCF and the Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS). Results show that snow DA improves temperature predictions especially in the Tibetan Plateau (by 5–20%) and high latitudes. Improvements at low latitudes are seen immediately and last up to 60 days, whereas improvements at high latitudes only appear later in transitional seasons. At high latitudes, assimilating GRACE data results in marked and prolonged improvements (by ~25%) due to large initial snow mass changes. This study has great implications for future land DA and seasonal climate prediction studies. (KEYWORDS: data assimilation; seasonal temperature prediction; MODIS; GRACE)

## 2.2. INTRODUCTION

The study of seasonal climate prediction beyond the time scale of weather forecasting requires seeking potential sources of predictability from the slowly varying ocean and land surface conditions. While sea surface temperatures (SSTs) have strong controls on temperature and

---

1 This chapter was published in Lin, P., J. Wei, Z.-L. Yang, Y.-F. Zhang, K. Zhang, 2016: Snow data assimilation-constrained land initialization improves seasonal temperature prediction. *Geophysical Research Letters*. DOI: 10.1002/2016GL070966. P.L. designed and performed experiments and analyzed results with advises from J.W. and Z.-L. Y. Y.Z. prepared the restart files from data assimilation runs, and K.Z. contributed to statistical analysis.

precipitation over certain tropical/subtropical regions (Shukla 1998; Shukla et al. 2000), land can memorize past climate information and contribute to atmospheric prediction over extended land areas and seasons (Guo et al. 2012; Koster et al. 2000a). Therefore, investigating land-derived seasonal climate predictability has received increased attention in recent years.

Realistic land initialization can contribute to seasonal climate prediction, as demonstrated by several modeling studies that focused on snow (Douville 2010; Jeong et al. 2013; Orsolini et al. 2013), soil moisture (Hirsch et al. 2014; Koster et al. 2004b; Koster et al. 2010b), vegetation (Koster and Walker 2015; Williams and Torn 2015; Xue et al. 2010), and groundwater (Jiang et al. 2009). To obtain reasonable estimates of initial land conditions, the most widely-used approach is running state-of-the-art land surface models (LSMs) in offline mode with observation-based meteorological forcing (Dirmeyer et al. 2006). For example, the Global Land–Atmosphere Coupling Experiment (GLACE-2) (Koster et al. 2011) and the aforementioned modeling studies adopted this approach to acquire an estimate of initial land conditions, which can be used in coupled land–atmosphere prediction. However, this approach could be problematic in deriving accurate initial land conditions due to deficiencies in model structure and parameterization and forcing errors. In particular, Koster et al. (2011) found that land-derived predictability is very limited in regions with sparse rain gauge observations, such as the Tibetan Plateau (TP) and high latitudes. These regions suffer from low quality of meteorological forcing data, and the forcing errors subsequently propagate into inaccurate initial land conditions, prohibiting useful information to be transferred to the coupled land–atmosphere prediction.

To address the issue with inaccurate land initialization, it is acknowledged that assimilating satellite data into offline LSMs could be useful to hydroclimate prediction (Koster et al. 2011; Koster et al. 2010a; Mahanama et al. 2012) as land data assimilation (DA) can combine the strengths of physically-based models and observations. However, in practice, very few studies have demonstrated the contribution of satellite-constrained land DA products in coupled land–atmosphere prediction. Many land DA studies were conducted at basin- or watershed-scales for hydrological applications such as streamflow prediction (Liu et al. 2015) and drought monitoring

(Kumar et al. 2014), but they were not used for climate prediction purposes. Lately, Santanello Jr et al. (2016) presented preliminary results on using soil moisture DA outputs, which assimilated Advanced Microwave Scanning Radiometer for EOS (AMSR-E) data, in short-term weather forecasting. Yet, the degree to which land DA can advance seasonal climate prediction is still largely unexplored, although the importance has been recognized for over a decade in the community (Santanello Jr et al. 2016). Reasons for a lack of the research efforts are associated with the complexity in conducting land DA especially at global scales, which needs to deal with land surface heterogeneity, a lack of land surface observations, and larger computational demand compared to open-loop (OL) LSM runs that do not assimilate land observations.

In this study, we present the first investigation to quantify the impacts of land initialization on seasonal prediction of surface air temperature in the Northern Hemisphere (NH) that emphasizes the role of two snow DA products generated from a recently developed global land DA system (Zhang et al. 2014). The system was developed based on the National Center for Atmospheric Research (NCAR) Data Assimilation Research Testbed (DART) and the Community Land Model version 4 (CLM4) (hereafter DART/CLM4). Zhang et al. (2014), Zhang (2015), and Zhang and Yang (2016) assimilated the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) and the Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS) into the offline DART/CLM4 to improve global estimates of snow, the products of which are used in the coupled modeling experiments presented in this study (see more details on the products in Section 2 and Section 3.1). Remote sensing-derived snowpack estimates by different satellite retrieval techniques, such as optical-band (Zhang et al. 2014), gravity-field (Su et al. 2010), and passive microwave (Durand et al. 2009; Kwon et al. 2015), each have unique strengths and limitations. Therefore, utilizing the snow DA products that assimilate MODIS SCF and GRACE TWS data, this study also investigates how the assimilation of optical-band and gravity-field satellite observations may influence seasonal climate prediction differently.

We focus on the three research questions. (1) Can NH snow DA help improve surface air temperature prediction at seasonal lead times? (2) How do the improvements vary across regions

and lead times, and why? (3) For coupled land–atmosphere prediction, what is the added value of assimilating GRACE TWS in addition to assimilating MODIS SCF only? Section 2 introduces the modeling system, experimental design, and datasets. Section 3 presents results and Section 4 closes with major conclusions and discussions.

## 2.3. MODELING SYSTEM AND EXPERIMENTAL DESIGN

We conducted three suites of ensemble-based experiments using the Community Earth System Model (CESM) version 1.2.1 for Atmospheric Model Intercomparison Project (AMIP)-type (Gates 1992) runs. SST and sea ice concentration were prescribed using the Hadley Centre monthly data (Hurrell et al. 2008). Therefore, the experiments were not true seasonal forecasts which initialize all components. Instead, the configuration was used to allow isolated investigations on land impacts (Koster et al. 2011). The CLM4 LSM was coupled with the Community Atmospheric Model version 5 (CAM5), and the model resolution was 0.9°×1.25°. CLM4 has up to five layers of snow depending on snowpack thickness, and the SCF parameterization is a function of snow density and snow depth (Niu and Yang 2007). CAM5 has 30 vertical layers.

**Table 2.1** summarizes the experimental design and model configurations, which only differed in the land initializations. Due to the data availability of snow DA results, each experiment covered seven years from 2003 to 2009, the overlapping time period for the MODIS and GRACE data. The experiments were initialized on January 1, February 1, and March 1 of each year; the analysis focused more on results initialized on March 1 because the impact of snow initialization in cold-to-warm season transitions is a relatively understudied issue (Thomas et al. 2016). There were 8 ensemble members for each start date and the length of each simulation was three months. In total, there were 504 three-month runs (3 suites×7 years×3 start dates×8-member ensembles). For atmospheric initialization, the ERA-Interim reanalysis data were interpolated to CAM5-consistent grids for initializing CAM5 3-D fields. To generate 8-member ensembles, the pre-

processed reanalysis data at the 00h of the ±4 days of each initialization date were used to perturb atmospheric initialization as similarly documented by previous studies (Saha et al. 2014). For land initialization, the restart files generated from the land DA system were used. The system used an ensemble Kalman filter to conduct DA (40-member), and the ensemble mean restart files were used to initialize CLM4. In Table 1, OL is the control experiment. MOD initializes land conditions based on the assimilation of daily MODIS SCF data. GRAMOD employs initial land conditions that jointly assimilate monthly GRACE TWS and daily MODIS SCF. Within each suite of experiments, the same land initialization file was used across the 8-mememberensembles. Readers can refer to Zhang et al. (2014), Zhang (2015), and Zhang and Yang (2016), and Supplementary Information for more details on the NH snow DA. It is beyond the scope of this study to discuss snow DA technical details. However, it is emphasized that the three sets of offline DART/CLM4 simulations went through the same hundred-year spinning-up. They were also driven by the same set of DART/CAM4 reanalysis forcing data that were bias-corrected using the Global Precipitation Climatology Project (GPCP) pentad precipitation and the Clouds and the Earth's Radiant Energy System (CERES) monthly radiation data. The processing ensured the differences among OL, MOD, and GRAMOD are only results of assimilating different satellite data.

| **Experiment** | **SST and sea ice concentration** | **Atmosphere (CAM5)** | **Land initialization (CLM4)** |
|---|---|---|---|
| **OL** | Prescribed using Hadley Centre monthly data [Hurrell et al., 2008] | Initialized using ERA-Interim data, 8-member ensemble | Offline CLM4 simulation without assimilating satellite data |
| **MOD** | | | Offline CLM4 simulation assimilating MODIS daily SCF |
| **GRAMOD** | | | Offline CLM4 simulation jointly assimilating MODIS daily SCF and GRACE monthly TWS |

**Table 2. 1.** Experimental design and model configuration.

We used the following reference datasets to assess the modeling results. (1) Daily snow depth data are from two sources: a gridded analysis that combines in-situ observations with an optimally-interpolated first-guess field from a simple snow model (compiled by the Canadian Meteorological Center (CMC)) (Brown and Brasnett 2010), and a gauge-based product from the Global Historical Climatology Network (GHCN, ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/). (2) Daily SCF data are from the retrievals by the MODIS/Terra (http://modis-snow-ice.gsfc.nasa.gov/?c=MOD10C2). (3) Daily 2-m air temperature ($T_{2m}$) data are also from two sources: the Hadley Centre gridded data ($T_{Hadley}$) compiled using observations over 27,000 stations (Caesar et al. 2006) and the ERA-Interim reanalysis ($T_{ERA\text{-}Interim}$), which is constrained by synoptic observations and has no missing value (Dee et al. 2011). $T_{Hadley}$ consists of daily maximum and minimum near-surface temperature data, and their averages were used to compare with model-predicted $T_{2m}$. These datasets represent the best available data for evaluating model results at the global scale. Despite the data source uncertainties, our general conclusions based on comparing OL, MOD, and GRAMOD are not influenced by the reference data used, while detailed interpretations should always take the data uncertainties into consideration.

## 2.4. RESULTS

### 2.4.1. Snow DA-induced differences in initial snow conditions

We first show the spatial pattern of initial snow conditions in the three suites of experiments, focusing on their differences between the DA cases (MOD, GRAMOD) and OL. Snow influences the atmosphere initially through snow–albedo feedback and then through soil moisture feedback after snow melts (Xu and Dirmeyer 2013). SCF (Fig. 2.1a) and snow depth (Fig. 2.1c) are important indicators of the albedo and the snow mass available for snowmelt, respectively.

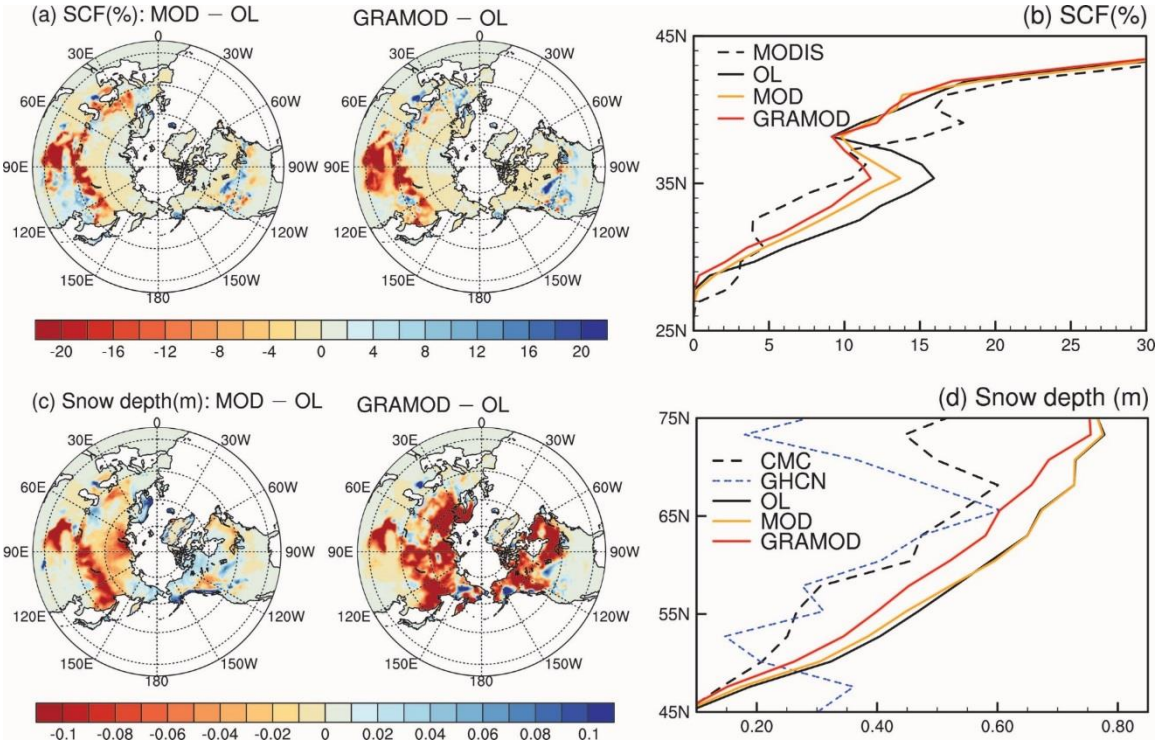


**Figure 2. 1.** Initial snow condition differences between DA and OL (left) and their zonal mean values compared with observations (right). (a) and (b) are for snow cover fraction in percentage; (c) and (d) are for snow depth in meters. The MODIS SCF data, the CMC snow depth analysis, and the GHCN gauge-based snow depth observation are used as reference datasets (dashed lines). The snow conditions averaged for March 1 of 2003 to 2009 are shown. The GHCN snow depth observation contains measurements from 9,908 stations for these dates.

Over a majority of regions, MOD and GRAMOD have a smaller initial snowpack than OL, suggesting that satellite DA reduces the OL-simulated SCF and snow depth in most areas, except for parts of Europe, North America, and North China. For SCF (Fig. 2.1a), the most significant differences lie in the TP and snow transition zones (middle latitudes between snow-covered and snow-free regions, 45–55°N), which is explained by the fact that the snow DA is most effective in regions with large snow uncertainty (Zhang et al. 2014). For snow depth (Fig. 2.1c), it is furthermore evident that GRAMOD reduces it much more effectively than MOD at high latitudes including northern Asia and northern North America. Since the retrieval algorithm of the MODIS SCF product is based on optical bands, the product is subject to two limitations: (1) it cannot detect additional snow when SCF reaches 100%, or it becomes "saturated", and (2) the data availability

is low at high latitudes during wintertime due to a lack of solar radiation. Thus, assimilating MODIS SCF data only brings limited changes in regions fully-covered by snow. By contrast, GRACE uses gravity-field observations and is able to measure snow mass variability even when SCF reaches 100%, providing much greater initial snow mass changes in GRAMOD than in MOD at high latitudes. Further comparisons based on zonal mean values (Figs. 2.1b and 2.1d) show that the DA cases are generally closer to the reference datasets including the MODIS SCF, the CMC analysis, and the GHCN observation. Note that all reference datasets contain uncertainties in certain regions due to gauge density and satellite retrieval limitations, thus zonal mean comparisons are provided to display the general effectiveness of DA across different latitudes. Fig. 2.1 does not necessarily indicate that DA is effective in improving initial snow estimates at all locations; instead, the independent evaluation on temperature predictions in Section 2.4.2 provides an alternative to assess the snow DA updates across a spatial scale.

### 2.4.2. Impact of snow DA-constrained land initialization on surface air temperature prediction

In Fig. 2.2, the cumulative root-mean-square-error (cRMSE) of the T2m prediction in OL, MOD, and GRAMOD are compared. cRMSE is calculated as:

$$\text{cRMSE} = \sum_{i=1}^{6} \sqrt{\frac{\sum_{y=2003}^{y=2009} [T_{model(i,y)} - T_{obs(i,y)}]^2}{n}} \quad (1)$$

where $T_{model}$ and $T_{obs}$ are model-predicted and observed T2m, respectively; n=7 years from 2003 to 2009; $i$ is the lead time binned every 15 days (hence 6 bins for a three-month forecast period). cRMSE accumulates the T2m prediction errors across the whole forecast period, and thus it is able to reduce possible noises as some noises can be cancelled out. In Figs. 2.5–2.7, additional assessments based on RMSE, square of correlation, and predictability (Koster et al. 2011) are also shown as a function of six lead time bins; such assessments are subject to more noises than cRMSE.

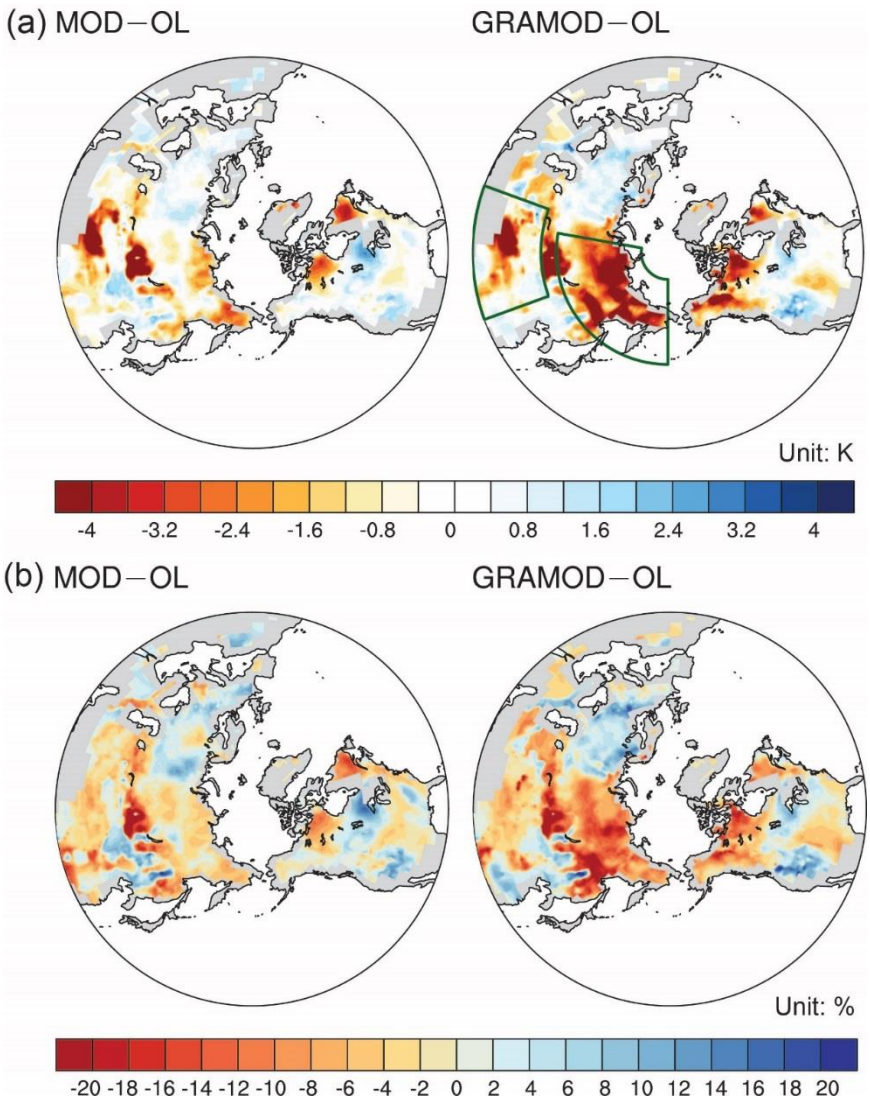


**Figure 2. 2**. The temperature prediction cumulative RMSE (cRMSE) difference between DA and OL. (a) is absolute value difference (K); (b) is percentage difference (%). Left (right) panel shows the difference between MOD (GRAMOD) and OL. The forecasts are initialized on March 1. Negative values indicate reduced prediction errors and improved temperature predictions after using snow DA-constrained land initializations. The green boxes encompass two regions of interest for a further analysis in Section 2.3.

Across the whole forecast period (March to May), MOD shows largest cRMSE reductions (up to 9 K) in the TP area and snow transition zones, and only slight cRMSE reductions at high latitudes. For GRAMOD, extended areas of the low-latitude regions are characterized by reduced cRMSE, while vast areas of high latitudes including Siberia, Alaska, and northern Canada observe

large cRMSE reductions (up to 11 K). Overall, the number of grid cells with improved $T_{2m}$ predictions accounts for 56.0% and 66.3% of the total grid cells in MOD and GRAMOD, respectively. In terms of percentage cRMSE change, the most prominent improvements for MOD and GRAMOD can be up to 30–40% regardless of reference datasets used. Over the TP, the percentage improvement is estimated to be only 5–8% for both MOD and GRAMOD using $T_{Hadley}$. However, the estimate can be more than 20% if using $T_{ERA\text{-}Interim}$ as the reference (not shown). Note that $T_{Hadley}$ has very low gauge density over the TP, which may artificially result in large bias and the subsequent low percentage improvement here. Over high latitudes such as Siberia, MOD brings less than 5% of improvement while GRAMOD can offer more than 25% of improvement. Further analysis taking different lead times into account (Figs. 2.5–2.7) reveals that, during the initial forecast period, the improved $T_{2m}$ predictions are mainly located at the TP and snow transition zones. At longer lead times, the improvements shift towards higher latitude regions. GARMOD generally exhibits better $T_{2m}$ predictions than MOD at high latitudes. $T_{2m}$ predictions are also degraded in some regions, which could be explained by three possible causes: (1) atmospheric noise due to the chaotic nature of the atmosphere, (2) sampling noise due to the limited number of samples used in computing statistics, and (3) possibly degraded snow DA results over particular boreal forest zones (Zhang et al. 2014).

In order to focus the analysis on local atmospheric responses directly forced by initial snow DA-induced differences, we plotted the RMSE differences between DA and OL (Fig. 2.3) by considering only regions with prominent initial snow water equivalent (SWE) differences (Fig. 2.8, regions with initial SWE differences less than 3 mm are masked out to reduce possible noise).

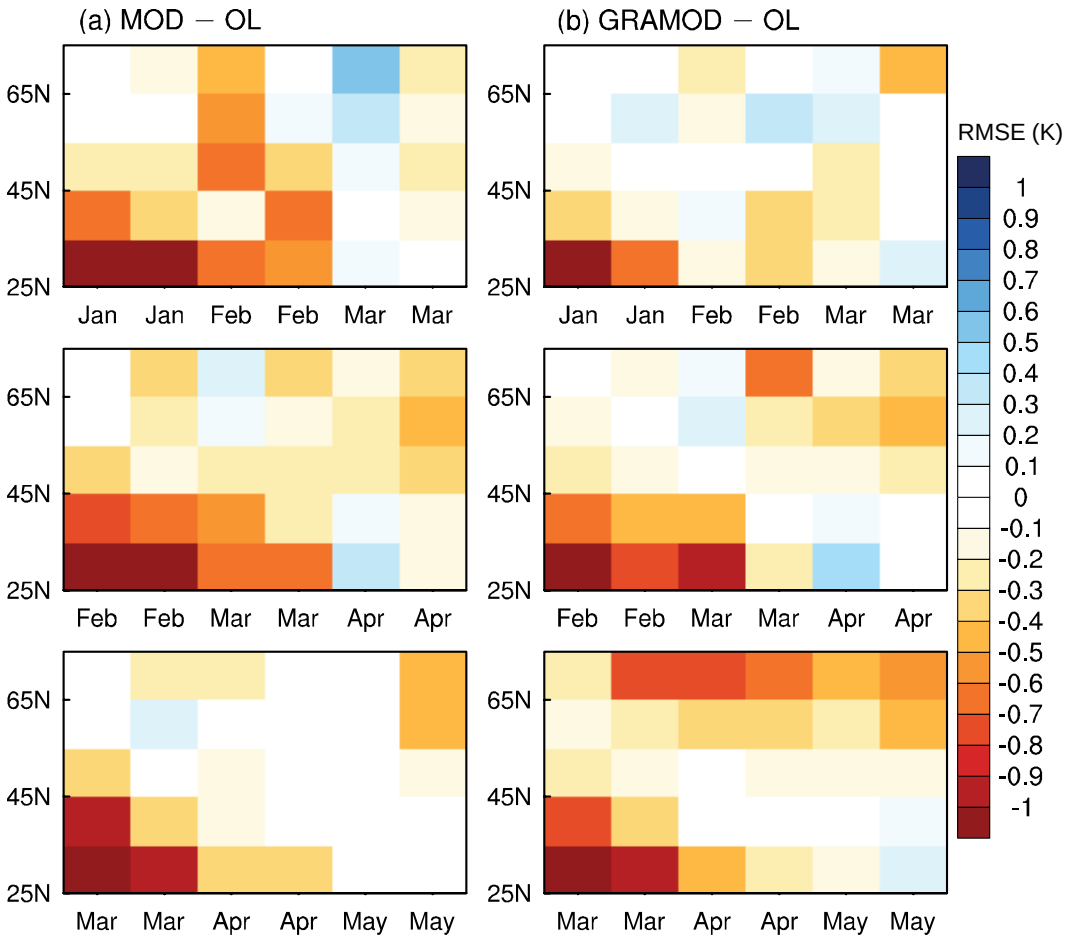


**Figure 2. 3**. The RMSE differences between MOD and OL (left) and GRAMOD and OL (right) in temperature predictions. X-axis is the lead time (binned every 15 days) and y-axis is the latitude (binned every 10 degrees from 25°N to 75°N). Negative RMSE differences indicate reduced prediction errors and improved temperature predictions. Model predictions initialized on January 1, February 1, and March 1 are shown in upper, middle, and lower panels, respectively. Regions with initial SWE differences greater than 3 mm are considered in the calculation.

In Fig. 2.3, reduced RMSEs are observed across different latitudes and lead times, suggesting that the initial SWE differences ranging from 3–30 mm (Fig. 2.8) generally lead to improved $T_{2m}$ predictions. At low to middle latitudes (25 to 45°N), snow DA can immediately improve $T_{2m}$ predictions regardless of start date. The improvements generally last up to 60 days, after which only marginal improvements are seen. In comparison, at middle to high latitudes (45 to 75°N), the improvements during the initial cold months are very limited, while pronounced improvements can be seen in warm months even at a long lead time of 75 to 90 days. The reasons

for the observed pattern are discussed in detail in Section 2.4.3. For short-lead forecasts initialized on January 1 (Figs. 2.3a and 2.3b, first row), greater and longer-last improvements at low latitudes are seen in MOD than in GRAMOD, which is related to the fact that MOD brings larger initial SCF changes in the TP on January 1 (see Fig. 2.4a, blue lines and current DA in handling MODIS and GRACE data inconsistency in Section 2.6). Since MOD shows better $T_{2m}$ predictions, it most likely suggests that the initial snow in MOD is better than that in GRAMOD in the TP. This has great implications for future multi-sensor snow DA studies to consider rule-based strategies to deal with data inconsistency. At higher latitudes, it is found that assimilating GRACE TWS in addition to MODIS SCF markedly improves the $T_{2m}$ predictions. The unique gravity-field observations by the GRACE TWS data are manifested in providing better initial snow mass estimates at high latitudes, addressing the "saturation" and data availability issues in the MODIS-constrained land initializations (Section 2.4.1). This is eventually translated into better $T_{2m}$ predictions.

### 2.4.3. A further regional analysis

To understand the latitude and lead time dependence of snow DA-induced improvements (Fig. 2.3), we focused on the TP and Siberia for a further analysis. These two regions display the best snow DA performances and $T_{2m}$ improvements. The TP is also a region characterized by strong snow-atmosphere coupling (Xu and Dirmeyer 2011). In addition to $T_{2m}$, we plotted the predicted time series of the area-averaged SCF, absorbed solar radiation (SA), and snow melt heat flux (SMH) over the two regions (Fig. 4). These four variables are direct model outputs of CLM4. SMH is the phase change energy associated with snowmelt calculated as the excess ground energy flux into the snow when melting criteria is met (Oleson et al. 2010).

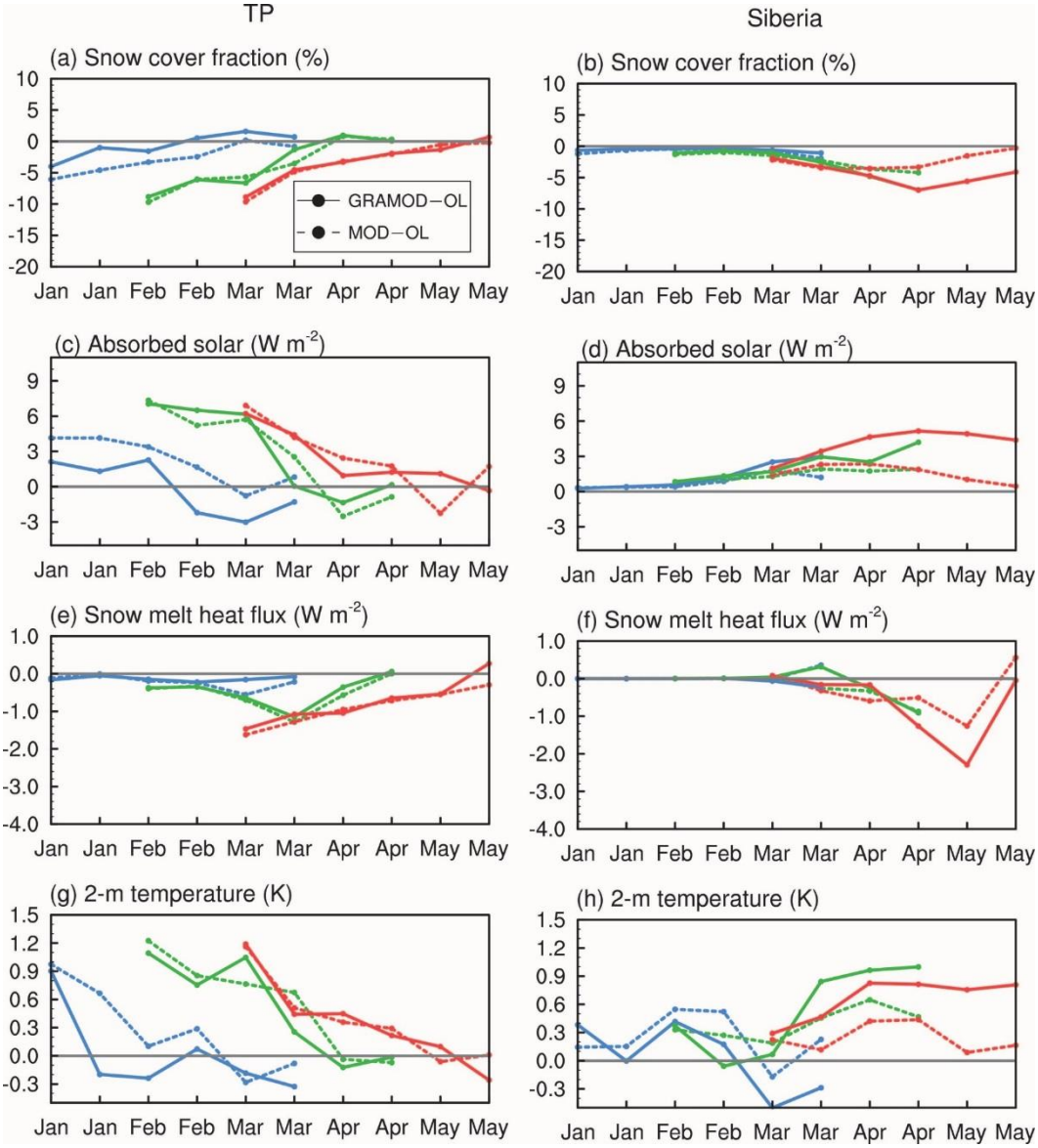


**Figure 2. 4**. Predicted time series of the differences between DA and OL in snow cover fraction (%), absorbed solar radiation (W $m^{-2}$), snow melt heat flux (W $m^{-2}$), and 2-m temperature (K). The values are 15-day mean averaged from 2003 to 2009. Model predictions initialized on January 1, February 1, and March 1 are shown in blue, green, and red colors, respectively. The TP (25–45°N, 70–110°E, above 2000 m) and Siberia (50–80°N, 80–180°E) are shown in left and right panels, respectively. These two regions are enclosed in the green boxes in Fig. 2.2.

Since snow DA reduces the initial snowpack estimates in both the TP and Siberia (Fig. 2.1), such regions would have more absorbed solar radiation, less snow melt, and more sensible heat transfer in DA cases than in OL, which together would lead to higher $T_{2m}$. Indeed, as shown in Fig.

2.4, the expected heat flux differences between DA and OL, i.e., positive SA differences and negative SMH differences, can be seen in both regions, although they are observed at different stages of a forecast.

The TP is a low-latitude high-elevation region, thus the incident solar radiation is strong from January to March (150 to 220 W $m^{-2}$). This strong incident solar radiation combined with the fact that snow DA reduces the initial SCF by 5% to 10% (Fig. 2.4a) immediately increases the SA over the TP. The large positive SA differences between DA and OL (+3 to +6 W $m^{-2}$) lead to positive T2m responses (+0.9 to +1.2 K) during initial forecast period, which gradually weakens as the snow and SA differences between DA and OL approach zero at long lead times. In comparison, in Siberia, the SCF and albedo differences prior to March are negligible (Fig. 2.4b) due to the nearly saturated SCF. Combined with weak incident solar radiation (0 to 100 W $m^{-2}$), the initial SA differences are small (+1.5 W $m^{-2}$) despite large initial SWE differences (up to –22 mm for GRAMOD). Therefore, regardless of the satellite data used for constraining initial land conditions in Siberia, snow DA plays only a marginal role in influencing $T_{2m}$ predictions in cold months. Entering April, Siberia starts to receive increasingly strong solar radiation, driving the positive SA differences and the negative SMH differences to peak. For GRAMOD, these peak values reach +5 W $m^{-2}$ and –2.5 W $m^{-2}$, respectively, leading to higher $T_{2m}$ which only becomes pronounced (+0.9 K) in warm months. This explains the delayed impact of snow DA on temperature prediction at high latitudes. In Siberia, GRACE DA-induced heat flux and $T_{2m}$ differences are still pronounced at the end of May (Fig. 2.4h). This suggests possible prolonged impacts of snow DA on $T_{2m}$ predictions even in early summer, warranting further investigation.

Using realistic soil moisture initializations, Guo et al. (2012) reported that the temperature and precipitation predictability in North America shows a rebound in spring-to-summer transitions when soil moisture–atmosphere coupling strengthens. Analogously, we find that snow DA-induced improvements in $T_{2m}$ predictions rebound in winter-to-spring transitions when snow–atmosphere coupling strengthens. This rebound is most obvious at high latitudes (Fig. 2.3), because during initial cold months the albedo effects are negligible, prohibiting the immediate contribution

of accurate land initialization from snow DA. Entering warm seasons, as elevated solar radiation drives both albedo and snowmelt effects, the impact of snow DA also reaches a maximum. The findings in this section imply that the regional and lead time dependence must be considered for snow DA to be effective in improving T2m predictions.

Finally, Fig. 2.9 shows that MOD and GRAMOD consistently reduce the prediction mean bias, which is especially helpful to regions that lack ground observations. In addition, GRAMOD consistently improves the anomaly prediction while the effect of MOD is small. Overall, snow DA shows promise for improving two important aspects of a seasonal prediction (i.e. anomaly and bias), but its contribution to the anomaly prediction needs to be further assessed in future studies by increasing the number of samples.

## 2.5. CONCLUSIONS AND DISCUSSION

This study is, to our knowledge, the first attempt to quantify how seasonal climate prediction can be advanced through the use of improved land initializations with snow DA, which assimilated the MODIS SCF and the GRACE TWS data. Overall, our results suggest that using snow DA-constrained land initializations can improve seasonal temperature prediction, most significantly in the TP and high-latitude regions. Without assimilating land observations, these regions are characterized by limited land-derived predictability (Koster et al. 2011) due to inadequate rain gauge observations which degrade initial land estimates.

The snow DA-induced improvements in $T_{2m}$ predictions depend on latitude and lead time. At low to middle latitudes, snow DA can improve $T_{2m}$ predictions immediately after a forecast begins regardless of start date, which generally lasts up to about 60 days. At high latitudes, the most pronounced $T_{2m}$ improvements are not seen during the initial forecast period in cold months, but rather in transitional months. The reason for the delayed improvement is related to the weak albedo effects during the cold months, and the combined albedo and snowmelt effects in warmer months driven by increased incident solar radiation.

Using different satellite-constrained snow DA products, we also conclude that assimilating GRACE TWS in addition to MODIS SCF offers marked additional improvements at high latitudes. The improvements induced by GRACE DA can be seen at a long lead time of 75–90 days and may even last to the summer, clearly highlighting the unique strengths of GRACE gravity-field observations in climate prediction studies. Recently, there has been increasing interests in developing integrated multi-sensor land DA systems for better constraining land surface variables (Lettenmaier et al. 2015). The DART/CLM4 system presented in this study is one such system, which has also been used to assimilate AMSR-E high- and low-frequency brightness temperature to improve large-scale snow and soil moisture estimates (Kwon et al. 2016; Zhao et al. 2016). However, we do not yet understand how observations can be optimized to improve the seasonal climate prediction skill considering the utility and limitations of various satellite data. To this end, the insights gained from this study could be helpful to both land DA and climate prediction communities. Our findings also have great implications for operational forecasting systems that integrate land DA in seasonal climate prediction, suggesting the need for future studies to be conducted in this direction.

## 2.6. SUPPORTING INFORMATION AND FIGURES

We summarize several key snow DA technical details here. The MODIS-only DA follows the scheme by Zhang et al. (2014), where the error of MODIS SCF is set to 0.1 based on previous studies. Snow water equivalent (SWE) is the state vector. The MODIS and GRACE DA uses two passes to conduct DA, which follows the scheme by Zhang (2015) and Zhang and Yang (2016). The first pass advances the model for one month and calculates the monthly mean TWS to compare to GRACE TWS. The second pass uses the monthly increment resulting from the first pass to update daily SWE and soil moisture. In the second pass, MODIS SCF is assimilated when available. The GRACE DA uses the CSR RL05 GRACE data (http://grace.jpl.nasa.gov/data/get-data/monthly-mass-grids-land/), where the error assumption is from the spatially varying error

product provided by the website. SWE and soil moisture are the state vectors. In the current snow DA scheme, no preference is made between the GRACE TWS and MODIS SCF data – they are assimilated sequentially into the system in the second pass. It is possible they may give opposite sign of increments in the state vectors; the sum of the increments was applied to the state vector.

**Figs. 2.5–2.7** show the RMSE, square of correlation coefficient, and potential predictability, respectively, as a function of lead time. For useful seasonal predictions, both the mean bias and the anomaly are important, although the anomaly is a relatively more important aspect to be captured because the model systematic bias is arguably easier to be corrected using post-processing techniques. Previous studies mostly focused on assessing square of correlation (Koster et al. 2011; Orsolini et al. 2013), which provides statistical information on the phase of the predicted anomaly. Here, our analysis focused more on RMSE because (1) our sample size is not big enough (7 years) for robust analyses based on correlation alone, (2) RMSE can be resolved into the mean bias and the anomaly components (Taylor 2001), and hence it can provide information on the magnitude of errors in both components, and (3) reducing the mean bias is important yet difficult particularly in data-sparse regions. The potential predictability ($r^2_{ideal}$, **Fig. 2.7**) follows the definition by Koster et al. (2011), in which the first ensemble member is taken as the synthetic "truth" and the average correlation coefficient between the remaining seven ensemble members and the first member is used to estimate the atmospheric component's practical predictability. The plot shows that there is increased $r^2_{ideal}$ that generally follows the latitudinal and lead time patterns that we reveal in the RMSE analysis. However, reduced $r^2_{ideal}$ is also widespread globally, which may be ascribed to the large sampling noises in our calculations. In summary, due to the limited sample sizes used, we focus more on RMSE analysis in our study to conclude the improved seasonal forecast, but apparently future research is needed to better understand the actual contribution of DA by overcoming the sample size difficulties in our analysis.

Although we show in **Fig. 2.9** that $\mathbf{r^2}$ is consistently improved in GRAMOD (but not in MOD), it should be noted that due to a lack of data samples, the skill seen here is not to comment on the anomaly correlation skill, which is the most measure important measure to indicate a

model's ability to make a prediction. To this end, future research is warranted to use longer time-period DA products to better quantify the skill improvement statistically for an improved understanding of how much snow DA indeed helps with the forecast skill. In addition, investigations only focusing on extreme snow years might also be of help.

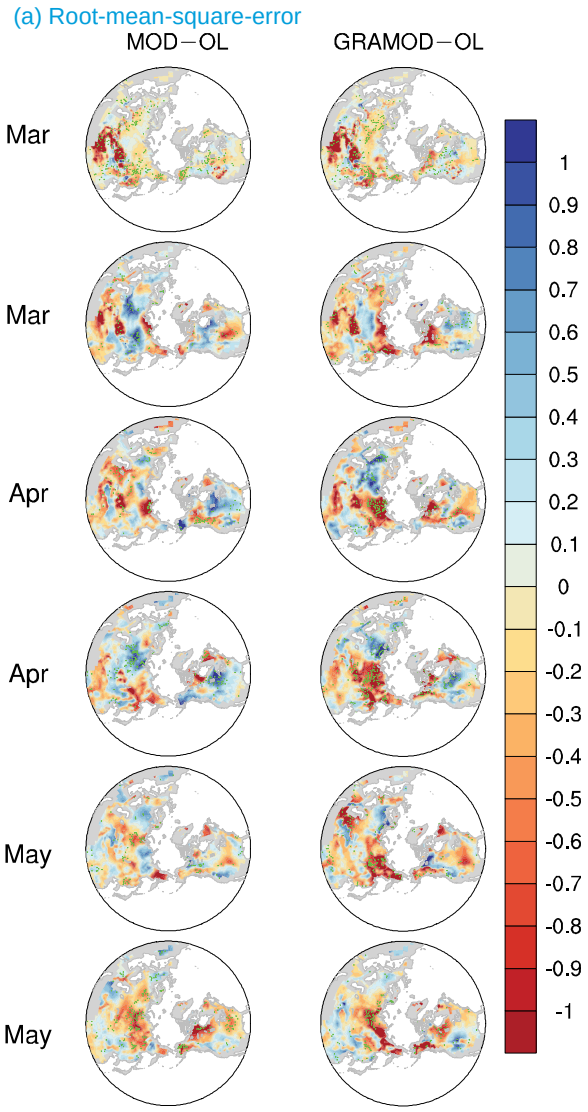


**Figure 2. 5**. The RMSE differences between DA and OL in temperature predictions as a function of lead time binned every 15 days. Left (right) panels show the results of MOD minus OL (GRAMOD minus OL). Green dots indicate the differences are significant at 95% confidence level according to a bootstrap method of randomly sampling the data for 1000 times.

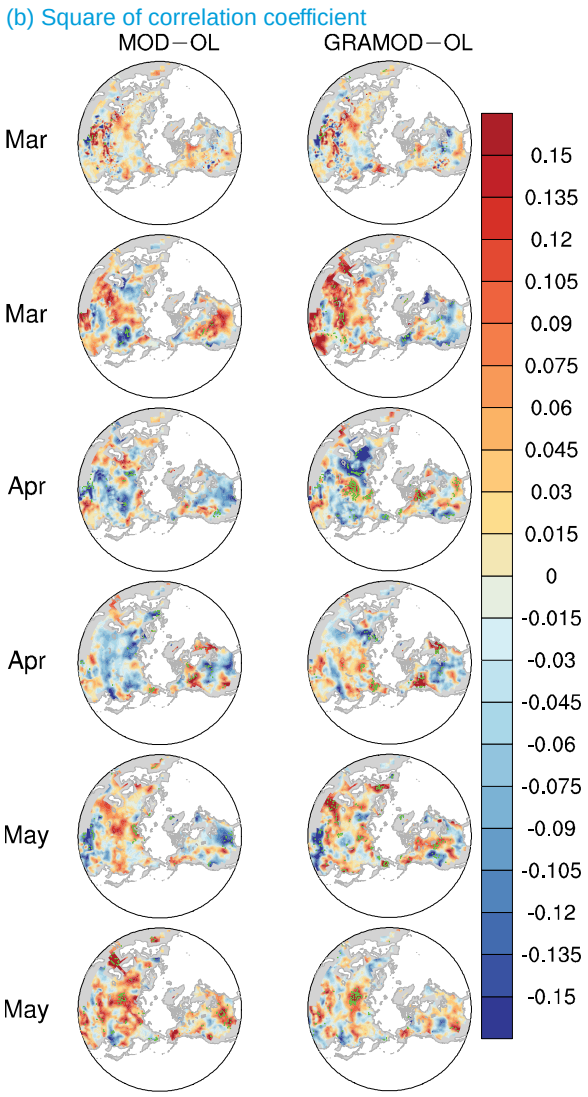


**Figure 2. 6**. Same as Fig. 2.5. but for the square of correlation coefficient ($r^2$), as defined by Koster et al. (2011).

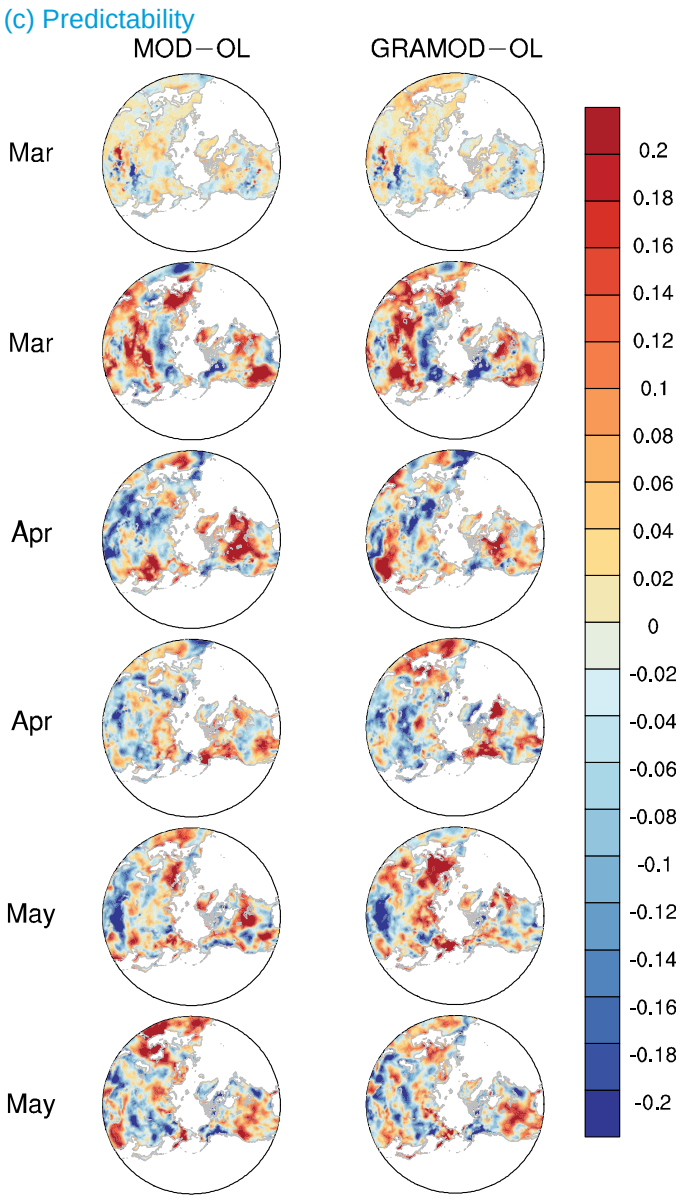


**Figure 2. 7**. Same as Fig. 2.5 but for potential predictability ($r^2_{ideal}$), as defined by Koster et al. (2011).

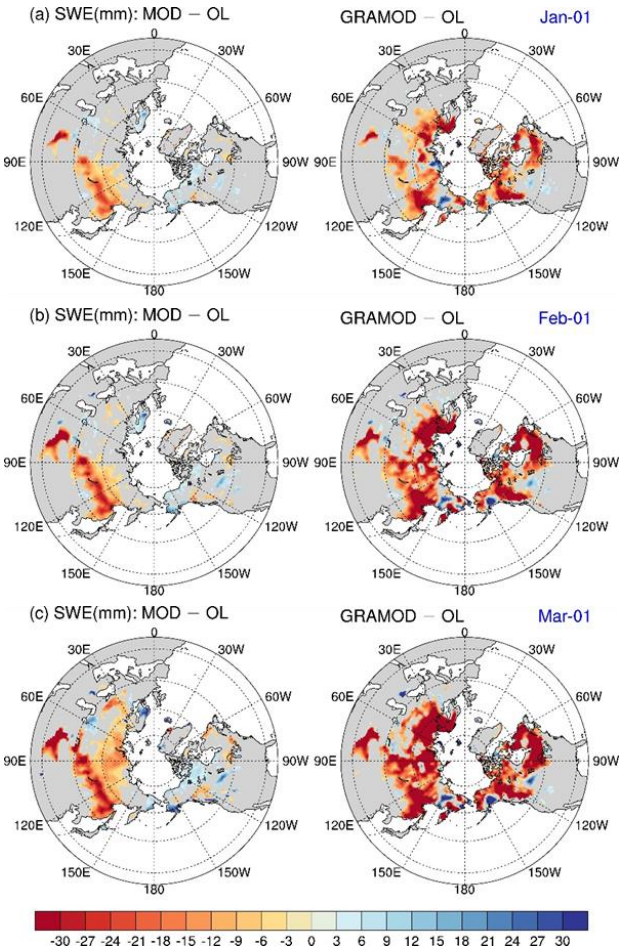


**Figure 2. 8**. Initial SWE differences as calculated using MOD minus OL (left) and GRAMOD minus OL (right). The initial snow differences for January 1, February 1, and March 1 averaged for 2003 to 2009 are shown in (a–c), respectively. Regions with initial SWE differences less than 3 mm are masked out.

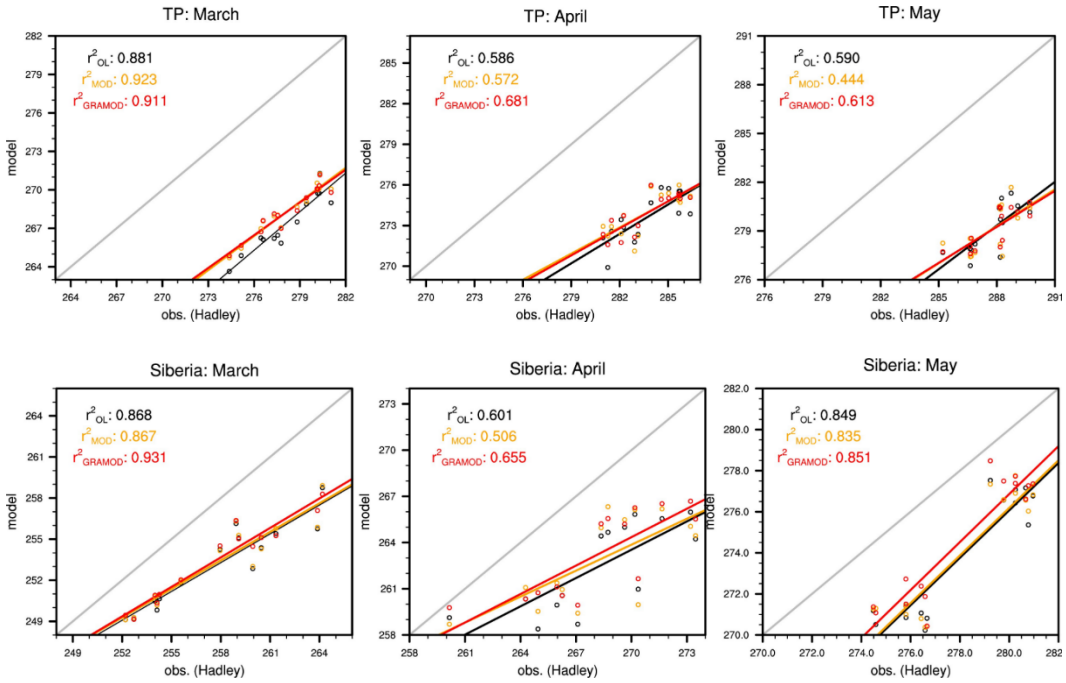


**Figure 2. 9**. Scatter plot between model-predicted $T_{2m}$ and observed $T_{2m}$ averaged over the TP (upper) and Siberia (lower). The prediction is initialized on March 1, and the scatter plot for March, April, and May is shown in left, middle, and right columns, respectively. Each plot contains 14 data samples (7 years×2 fifteen-day bins). The regression lines and the square of correlation coefficient are shown for OL (black), MOD (orange), and GRAMOD (red). Grey line is the 1:1 line.

## 2.7. ACKNOWLEDGEMENT

This work is funded in part by the National Natural Science Foundation of China under Grant 9133721 and in part by the Jackson School of Geosciences. We would like to thank Jerry Olson (National Center for Atmospheric Research) for sharing a package that allows the interpolation of reanalysis data to CAM5-consistent grids. Shaikh J. Muhammad and Long Zhao are thanked for technical support in running CESM experiments. Adam S. Papendieck is acknowledged for English editing and enhancing the quality of the manuscript. We would also like to thank two anonymous reviewers and Randal Koster for making suggestions to improve the manuscript. The data used in this study are available upon request to the corresponding author (liang@jsg.utexas.edu).

# Chapter 3: Skill of Asian monsoon seasonal forecast derived from springtime snow data assimilation

## 3.1. ABSTRACT

Skillful seasonal forecast of the Asian monsoon is an unresolved challenge for dynamical climate models that primarily relied on the memory of tropical oceans. Here we show evidence that additional prediction skill can be derived from improved springtime snow initializations by assimilating the Moderate Spectral Imaging Satellite (MODIS) and the Gravity Recovery and Climate Experiment (GRACE) data. Data assimilation (DA) provides significant snow updates over the western Tibetan Plateau (TP) and mid- to high-latitude Eurasia (EA). Together, they modulate the monsoon circulation by thermally acting on the Eurasian landmass at different times of the year. Then, a more robust rainfall forecast improvement can be seen over central north India (CNI) than other Asian monsoon subcomponents. For the CNI rainfall forecast in spring, the TP snow is more important, and MODIS DA slightly outperforms that with joint GRACE DA. For the forecast in summer (3-month lead), the additional skill mainly comes from the EA snow with GRACE DA, highlighting the geographic regions where DA needs to be refined for tailored Asian monsoon forecast. This study clarifies the complexities needed for successfully translating DA to useful monsoon prediction skill, which is important for seeking sources of monsoon predictability from future multi-sensor snow DA methodologies. (Keywords: Asian monsoon seasonal forecast; Snow data assimilation; Land-atmosphere interactions; GRACE; MODIS; Indian monsoon)

## 3.2. INTRODUCTION

The Asian monsoon affects more than sixty percent of the world's population (Wu et al. 2012), yet its skillful seasonal forecast has been an unresolved challenge (Doblas-Reyes et al. 2013; Webster et al. 1998). To obtain a skillful monsoon forecast, state-of-the-art dynamical climate models primarily rely on the memory of tropical oceans (Goddard et al. 2001; Palmer and Anderson 1994) and data assimilation (DA) methodologies to improve ocean initializations (Chen

et al. 1995). Land contains a memory of weeks to months (Koster et al. 2004b) that can potentially advance the monsoon forecast skill, however, it is much less constrained by observations in dynamical climate models than the atmospheric and ocean components (Balmaseda and Anderson 2009; Lahoz and Schneider 2014). Despite increased satellites monitoring snow, soil moisture, and vegetation that can characterize global land conditions together with physically-based land surface models (LSMs) through DA (Koster et al. 2010a; Mahanama et al. 2012; Rodell et al. 2004), in practice, these land-oriented satellite data are still underutilized in dynamical climate predictions. Our understanding is therefore greatly limited as to the additional skill that can be derived from land DA (Lin et al. 2016; Santanello Jr et al. 2016) and whether the intrinsic limit of seasonal climate predictability has been reached (Koster et al. 2004b).

Eurasian snow has long been recognized as a key for the Asian monsoon predictability (Blanford 1884). Typically, positive (negative) snow anomalies over the Eurasian landmass tend to be followed by negative (positive) Asian monsoon anomalies in the absence of other forcings (Senan et al. 2016). To obtain a robust snow estimation in a statistically optimal way, numerous research studies have developed snow DA methodologies (Slater and Clark 2006; Su et al. 2010; Sun et al. 2004), with their main goals to improve point- to basin-scale hydrological predictions. Very few snow DA studies have been conducted at the global scale (Kwon et al. 2016; Zhang et al. 2014) to address problems with seasonal climate prediction using general circulation models (GCMs), including the monsoon forecast. Widely-used global climate reanalyses such as the ERA-Interim and the Global Land Data Assimilation System (GLDAS) have initial data assimilation capabilities to ingest station-based snow depth and satellite-derived snow cover data (Dee et al. 2011; Rodell et al. 2004), which could be potentially used to satisfy the research goal. However, signals from snow DA cannot be separated from these reanalyses, thus our understanding of their effectiveness in atmospheric predictability is constrained.

Recently, Zhang et al. (2014) and Zhang and Yang (2016) assimilated the Moderate Resolution Imaging Spectroradiometer (MODIS) snow cover fraction (SCF) and the Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage (TWS) data into a state-of-

the-art land surface model (LSM) – the Community Land Model 4 (CLM4), which generated two global snow DA products for use in GCMs (hereafter Z1416). Although limited, they offer new opportunities to carefully quantify the role of snow DA in seasonal climate prediction. Lin et al. (2016) utilized them and demonstrated an improved temperature forecast of 5%–25% locally over regions with significant DA updates, which reveals interesting spatiotemporal patterns for the DA effectiveness. However, forecasting precipitation is a more compelling prediction problem. To the best of our knowledge, no existing studies have been attempted to explore monsoonal rainfall prediction from the perspective of snow DA.

The goal of this study is therefore to quantify the influence of DA-constrained springtime snow initialization on seasonal precipitation forecast over the Asian monsoon area using a series of coupled land–atmosphere "hindcast" experiments (see Section 3.2 and **Table 3.1** for details). Because the methodological development of global snow DA is still an area of active research (e.g. Sun et al. 2004; Zhang et al. 2014), we also focus on regions exhibiting the most robust skill improvement to assess the examined DA methods, aiming to provide discussions on future DA refinements for the tailored Asian monsoon prediction. Complexities with translating snow DA to useful monsoon prediction skill are also clarified, as are the roles of multi-sensor observations. In this study, we focus on the springtime initialization (1 March on 2003 to 2009) because the role of snow is relatively less understood during the cold-to-warm season transitions (Lin et al. 2016; Thomas et al. 2016). It is also a critical time span where sea surface temperature (SST) offers limited predictability to the atmosphere due to the so-called "spring barrier" in the El Nino Southern Oscillation (ENSO) prediction (Webster and Hoyos 2010). Thus, seeking sources of predictability besides that of SSTs is of particular interest during this transitional time span.

## 3.3. METHODS

### 3.3.1. Models and experimental design

The National Center for Atmospheric Research (NCAR) Community Earth System Model (CESM v1.2.1) is used to perform the coupled land–atmosphere "hindcast" simulations. Each suite of experiments is initialized on 1 March of 2003 to 2009 (due to the data availability of the snow DA products), and is integrated for six months. SSTs and sea ice are prescribed using the Hadley Centre monthly data (Hurrell et al. 2008); thus, the experiments do not represent actual dynamical forecasts that initialize all model components, but are instead conducted to provide an isolated investigation on the land impact. The Community Atmosphere Model (CAM5) and Community Land Model (CLM4) are used, because CLM4 is the host land model for snow DA with Z1416, and CAM5 is the most updated atmospheric model in CESM v1.2.1. The model resolution is 0.9°×1.25°. Z1416 produced 40-member ensemble land model outputs, and the ensemble average is taken as the land initial condition. Note that Z1416 used the Ensemble Adjustment Filter (EAKF) to conduct DA, which requires estimating the "prior" error distribution for the model states. To satisfy such a requirement for ensemble-based DA, Zhang et al. (2014) used 40-member bias-corrected forcing to introduce the model uncertainty, which resulted in the 40-member DA outputs.

To account for atmospheric chaos in the forecast, the atmospheric component (CAM5) also needs to be properly initialized and then perturbed. To satisfy this purpose, we pre-processed the time-shifted ERA-Interim data (±4 days of 0000 UTC 1 March) to be consistent with CAM5 to create an 8-member ensemble; such techniques to perturb the atmospheric initialization are commonly used in previous seasonal climate prediction studies (Saha et al. 2014). Across the 8 ensemble members, the same land initialization is consistently used. Readers can find more model configuration details in Lin et al. (2016).

**Table 3.1** summarizes the experiments conducted in this study. The first three experiments are designed to investigate the role of DA with respect to the utility of different satellite observations (i.e. MODIS and GRACE). The last two experiments are designed to understand the

relative importance of the TP and EA snow for tailored Asian monsoon seasonal forecast. There are in total 280 six-month modeling experiments (5 suites × 8-member ensembles × 7 years) for quantifying the role of snow DA in the Asian monsoon seasonal forecast.

| | **Land Initialization** |
|---|---|
| **The role of multi-sensor DA:** | |
| **OL** | "Open-loop" experiment: land initialization without assimilating satellite data |
| **MOD** | "MODIS" DA experiment: land initialization with global assimilation of the MODIS SCF data (Zhang et al. 2014) |
| **GRAMOD** | "GRAMOD and MODIS" DA experiment: land initialization with joint assimilation of the global GRACE TWS and MODIS SCF data (Zhang and Yang 2016) |
| **The role of regional DA (Fig. 3.12):** | |
| **TP_only (GRAMOD)** | Land initialization by applying GRAMOD over the TP only (Fig. 3.1); other regions' land initialization is the same as OL |
| **EA_only (GRAMOD)** | Land initialization by applying GRAMOD over the EA only (Fig. 3.1); other regions' land initialization is the same as OL |

**Table 3. 1.** Experimental design for ensemble-based coupled land–atmosphere experiments. Model is initialized on 1 March of 2003 to 2009, and is integrated for six months till the end of August each year. SST and sea ice are prescribed using the Hadley Centre data. The time-shifted ERA-Interim data are used to create 8-member ensemble for the atmospheric initialization. See more details of the model configurations in Methods and Lin et al. (2016).

### 3.3.2. Regional DA experiments

This study also applies GRAMOD DA over certain regions to understand which regional importance of DA. To apply DA constraints to a specified region only, we modified the 40-member ensemble mean land restart files, in which the columns, plant functional types (PFTs), and land units outside the predefined geographic region are replaced with those from the OL restart file. To ensure other land state variables are physically-consistent with the initial snow conditions without

imposing model imbalance errors, all snow-related variables for water balance, and soil and vegetation temperature-related variables for energy balance are also replaced with OL accordingly outside the specified region. The initial snow depth difference (m) between regional DA and OL is therefore shown in **Fig. 3.12**, and the geographic bounding boxes are shown in **Fig. 3.1**. The regional DA restart file is then used to initialize the land component of the ensemble-based "hindcast" experiments as described above.

### 3.3.3. Statistical analyses and reference datasets

Due to the limited available snow DA products (7 years from 2003 to 2009), we used the following skill metrics ($r^2$ and RMSE), bootstrap statistical tests, and five reference precipitation datasets to ensure robustness of the detected DA signal from the noisy atmospheric background. Below, equations for these skill metrics are shown for one reference precipitation; the same equations are used when comparing model outputs against other reference datasets.

For each season in the forecast period, $r_i^2$ for each ensemble member is first computed at each grid cell using 21 data samples (3 months×7 yrs) from the model and the observation, denoted as $\boldsymbol{P_m}$ and $\boldsymbol{P_o}$ vectors, respectively (Eqn. 3.1); $\delta\mathbf{P_m}$ and $\delta\boldsymbol{P_o}$ denote standard deviation and $cov$ stands for the covariance. $r^2$ is then calculated by averaging $r_i^2$ for the 8 ensemble members (Eqn. 3.2; see Section 3.3.1 for the ensemble generation technique), which is eventually used to compute the skill difference between DA and OL using Eqn. (3.3). Negative correlation is set to zero before the square operation in Eqn. (3.1) to reduce potential sampling noise (Koster et al. 2010b) although we found the influence of this procedure is minimal to our results.

$$r_i^2 = \left(\frac{cov(\boldsymbol{P_m}, \boldsymbol{P_o})}{\delta\mathbf{P_m} \cdot \delta\boldsymbol{P_o}}\right)^2 \qquad \text{Eqn. (3.1)}$$

$$r^2 = \frac{1}{n}\left(\sum_{i=1}^{i=n} r_i^2\right) \qquad \text{Eqn. (3.2)}$$

$$r_{diff}^2 = r_{DA}^2 - r_{OL}^2 \qquad \text{Eqn. (3.3)}$$

To test the null hypothesis that $r^2_{diff}$ is not significantly different from zero, at each grid cell, a 1,000-member bootstrapping procedure is applied to randomly resample the 21 values in the $\boldsymbol{P_m}$ vector in Eqn. (3.1). This procedure leads to 1,000 $r^2_{diff,shuffle}$ values with Eqns. (3.2–3.3). The null hypothesis is then rejected when the percentage of $r^2_{diff,shuffle}$ (denoted as α) is exceeded by $r^2_{diff}$ with a two-tailed distribution. Since we used 21 data samples in the observation and model data vectors (3 months and 7 years), the skill presented here incorporates both interannual and intraseasonal variance, where the magnitude of $r^2$ represents the explained fraction of total variance. Thus, interpretations for the skill presented in this study should differ from those using the conventional anomaly correlation skill.

RMSE is also assessed as another quantitative measure of the prediction (see Eqn. 3.4; N=21 for sample size; n=8 for ensemble size). Compared to $r^2$, RMSE can be more clouded with biases in the model and the observation, which explains its limited use in previous studies to detect land–derived prediction skill (e.g. Koster et al. 2011). However, DelSole and Shukla (2010) showed that the model bias and the model skill are interrelated, which necessitates reduced RMSE as another desired metrics for more comprehensive understanding of a seasonal forecast. Therefore, RMSE is jointly used in this study together with $r^2$. To reduce noise of the raw RMSE field, five reference datasets are used in computing RMSE (Eqns. 3.4–3.5), and only the consensus percentage RMSE (PRMSE) is presented in **Fig. 3.9**, in which we define a consensus only when the signs of $PRMSE_{diff}$ are agreed among ≥ 4 references. Once a consensus is reached, $PRMSE_{diff}$ are averaged among the agreed observations as the final result, and regions with annual climatological precipitation ≤ 200 mm are masked (**Fig. 3.9**).

$$RMSE_s = \frac{1}{n}\left(\sum_{i=1}^{i=n}\left[\sqrt{\frac{\sum_{j=1}^{j=N}\left(P_m^j - P_o^j\right)^2}{N}}\right]\right) \quad \text{Eqn. (3.4)}$$

$$PRMSE_{diff} = (RMSE_{DA} - RMSE_{OL})/RMSE_{OL} \quad \text{Eqn. (3.5)}$$

The five reference precipitation datasets include monthly precipitation data from the Global Precipitation Climatology Project (GPCP) and Center (GPCC), the Tropical Rainfall Measuring Mission (TRMM), the University of Delaware gridded precipitation (UDEL), and the National Oceanic and Atmospheric Administration (NOAA) Precipitation Reconstruction over Land (PRECL). All datasets are preprocessed to the model resolution (0.9°×1.25°) before assessment.

## 3.4.RESULTS AND DISCUSSION

### 3.4.1. Role of global snow DA

We are first interested in showing which geographic regions feature the most pronounced DA updates. Apparently, satellite observations only significantly modify the LSM-simulated snow over certain regions of the globe. **Fig. 3.1** shows the initial snow water equivalent (SWE) differences between the OL and DA experiments (see **Table 3.1**, where OL stands for "open-loop" without DA, MOD for experiment with MODIS assimilation, and GRAMOD for experiment with joint MODIS and GRACE assimilation). Most prominently, DA reduces the OL-simulated snow over majority of land areas (red colors with negative values in **Fig. 3.1a**), which is in line with alleviating the problem of the high-latitude positive snow biases in CLM4 (Toure et al. 2016) and the TP cold biases in most models (Su et al. 2013) reported by previous studies. Exceptions of increased snow are seen in parts of the Rocky Mountains and the Alaska (blue colors), which is geographically confined and less prominent. The globally-averaged magnitude of difference (bar charts in **Fig. 3.1a**) reveals that DA is more effective in certain years than others, such as 2004–2005 and 2008–2009 (see **Fig. 3.6** for the individual changes in snow cover fraction and snow depth, respectively). This suggests that DA-constrained snow initialization has year-to-year variations, which results from the DA mathematics that dynamically weights the relative uncertainty between the model and the observation.

GRACE observations are unique as they measure mass changes over time and are able to provide snow mass information that is absent in optical sensor observations, such as the MODIS snow cover fraction (SCF) observations. As a result, GRAMOD DA offers more prominent updates to the high-latitude snow depth (SD, **Fig. 3.1b**, right). However, MODIS DA is more effective in reducing SCF than GRAMOD DA over the TP (**Fig. 3.1c**, left), which suggests that MODIS and GRACE are inconsistent when updating the TP SCF (see a sensor-dependent behavior will be discussed more in Section 3.4.4). Zhang and Yang (2016) assessed the uncertainty with Z1416, reporting remaining issues with MODIS DA over the boreal forests (Zhang et al. 2014) and GRACE DA over certain river basins that have large surface water variations but were not yet accounted for when they disaggregate TWS signal in their DA methods (Zhang and Yang 2016). Nevertheless, the overall effectiveness of Z1416 was demonstrated (Zhang and Yang 2016) and further confirmed by Lin et al. (2016) in a coupled modeling experiment focusing on temperature. We note that dealing with land surface heterogeneity still presents the most challenges remaining to be resolved for global land DA studies, however, Z1416 is used in this study because our goal is to assess how much additional skill it may bring on top of that without DA, in order to guide the initialization strategies beyond what is currently adopted by most forecasting centers.

Overall, regions with prominent DA updates include the western TP, snowline areas around 50°N–55°N, and the higher latitudes (**Fig. 3.1**). These snow-dominated areas feature large model uncertainty due to their high seasonal snow variations (for snowline regions) and sparse rain gauges to serve as meteorological inputs and accurately force the model (for the TP and high latitudes). Because of the large uncertainty in their snow states, these regions are more prone to ingesting satellite observations than other regions of the globe because DA tends to give more weight to the observation when calculating the "posterior", if the modeled snow states suffer from large uncertainties. We next focus our analyses on the western TP and the mid- to high-latitude Eurasia (hereafter TP and EA, shown by green boxes in **Fig. 3.1**), because snow anomalies over these regions are key to the Asian monsoon predictability (e.g. Turner and Slingo 2011).

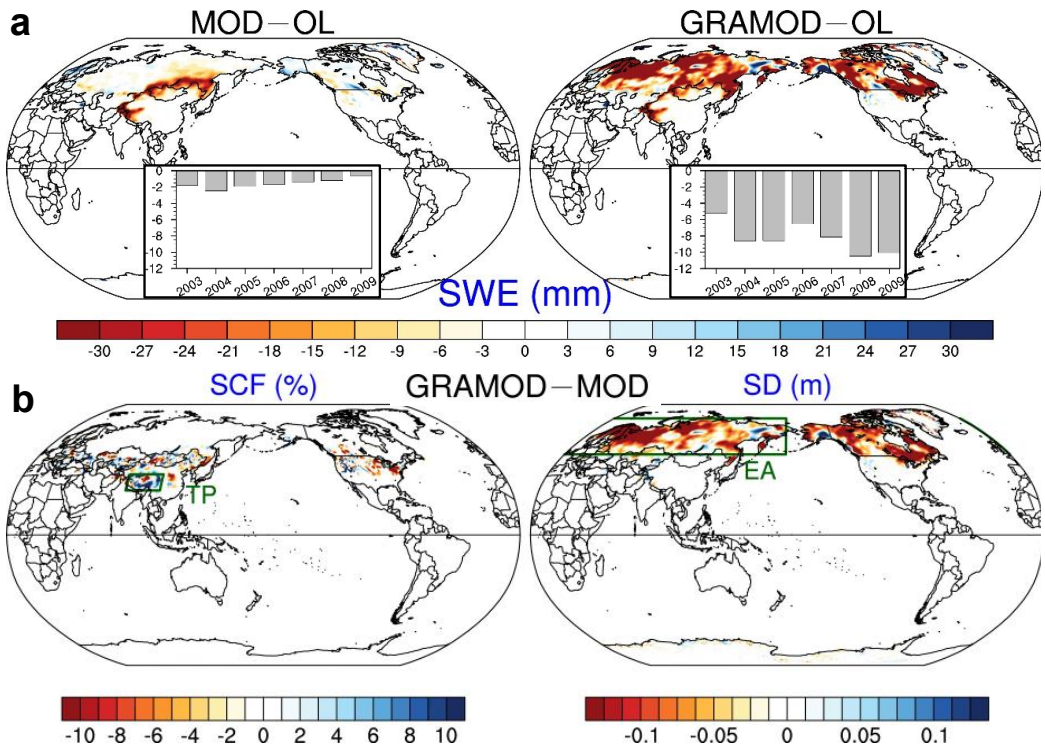


**Figure 3. 1.** DA-introduced changes in (a) initial snow water equivalent (SWE, mm) and (b) GRACE DA-introduced changes in initial snow cover fraction (SCF, %) and snow depth (SD, m). Calculation is performed using 7-yr averages on 1 March 2003 to 2009; bar charts show the year-to-year variation averaged over the globe.

### 3.4.2. Thermodynamic changes forced by snow DA

To facilitate physical understanding, DA-introduced initial SWE changes are viewed as external forcing brought into the system, where the forcing persistence is examined in the forecast months. With reduced snow by DA (negative $SWE_{diff}$ in **Fig. 3.2a**, first row), more absorbed solar radiation from reduced albedo and less latent cooling from reduced warm-season snow melt are expected. Hence, the thermal effect of DA is a positive sensible heating (**Fig. 3.2b**, second row) that warms up the overlying atmosphere locally over the TP and the EA. During March and April, the TP warming dominates the DA signal (**Fig. 3.2b**, blue lines) because it receives strong incident solar radiation that can translate any snow changes into immediate atmospheric warming (Lin et al. 2016). During May and June as EA receives more solar radiation to establish stronger snow-

atmosphere coupling (Xu and Dirmeyer 2011), the EA warming starts to dominate the DA signal (**Fig. 3.2b**, red lines). The EA signal can last longer into summer months in GRAMOD due to the long memory with the thick snowpack particularly over western Siberia before all snow DA signals vanish by the end of July (see **Fig. 3.7** for the surface air temperature responses). Therefore, although snow DA is applied globally in the springtime, it primarily operates on different locations of the Eurasian landmass, forcing the atmospheric warming at different times of the year to influence land-sea thermal contrast in different forecast months.

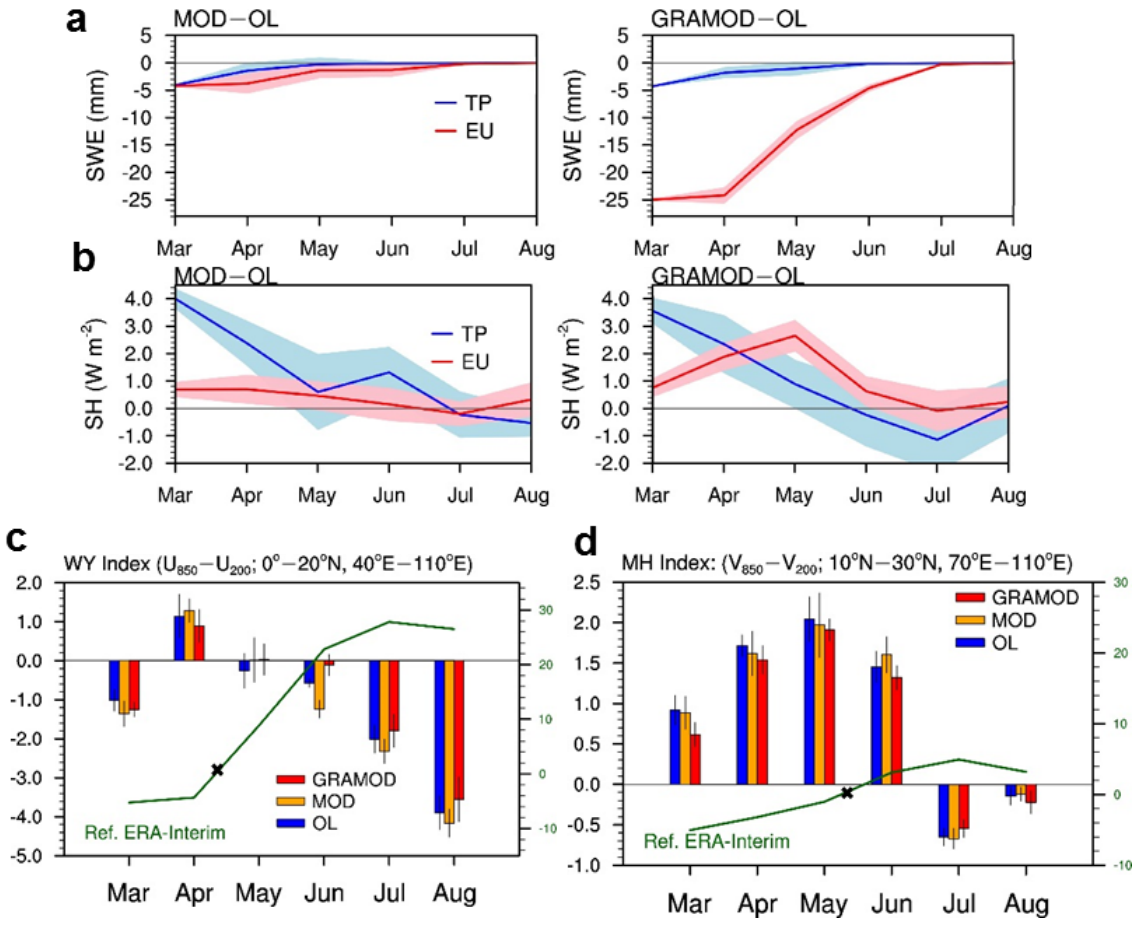


**Figure 3. 2.** Persistence of DA-introduced forcing and its influence to the monsoon circulation. (a) (b) show persistence of the SWE changes and the sensible heating (SH) respectively. (c) (d) show model-predicted monsoon circulation indices as biased against ERA-Interim (zero indicates no bias; cross marks the time when the circulation shifts its direction). Vertical lines and shading denote one standard-deviation uncertainty. The WY index (Webster and Yang 1992) and MH index

(Goswami et al. 1999) respectively summarize the Asian monsoon broad-scale circulation in zonal and meridional directions.

We next focus on analyzing the monsoon circulation changes in the DA experiments, because the thermal forcing brought by snow DA is expected to influence the land–sea thermal contrast, which is the most fundamental driver of the monsoon circulation (Chou 2003) among possible thermodynamics and dynamics mechanisms that can potentially communicate the snow DA signal to the monsoon area. To simplify the characterization of the highly complex Asian monsoon dynamics, we focus on two most widely used broad-scale monsoon circulation indices (**Figs. 2c–2d**) – one is the WY index (Webster and Yang, 1992; defined as the zonal wind shear between U850 and U200) and the other is the monsoon Hadley circulation index (MH index, Goswami et al. 1999; defined as the meridional wind shear between V850 and V200). In both the zonal and meridional directions, DA generally has less biased monsoon circulation indices than the OL cases. GRAMOD is more effective in reducing the monsoon circulation biases than MOD, and the associated uncertainty is the smallest during May and June, overlapping with the time when the EA signal dominates and the monsoon circulation shifts its direction for the onset (**Figs. 2c–2d**). The contribution of MOD is more consistently seen in the meridional circulation (i.e. MH index) than in the zonal circulation (i.e. WY index), which is likely because the MOD signal is geographically confined only to the TP region. However, due to the important thermal control of the elevated TP in influencing the Hadley circulation (Wu et al. 2012; Wu et al. 2007), the additional 4.8 W $m^{-2}$ sensible heating (representing 5%–6% of the TP springtime sensible heating) can exert controls on the meridional monsoon circulation albeit the signal decreases into May (**Fig. 3.2b**). In June and July, GRAMOD becomes even more effective in reducing the circulation biases than May and June (when the most obvious surface heating occurs), which may be ascribed to the positive feedback associated with the latent heat release aloft the Indian subcontinent (see **Fig. 3.8** for the warming and wind directions), which strengthened monsoon circulation in GRAMOD. This mid-tropospheric warming associated with the convective condensation is known to play a more important role in controlling the monsoon circulation strength than surface land-sea thermal

contrast (Dai et al. 2013), suggesting a positive feedback in GRAMOD that may make it contribute more to the Asian monsoon seasonal forecast.

### 3.4.3. Robust improvement in rainfall prediction over central north India

We then quantify the forecast skill that is derived from snow DA through differencing the skill between DA and OL, using the Global Precipitation Climatology Project (GPCP) monthly precipitation as the reference (see Methods). The skill is additionally assessed using four other precipitation datasets and the root-mean-square-error (RMSE) metrics (see Methods); due to the general agreement of the results in the spatial pattern (see **Fig. 3.9** for the consensus skill), only $r^2$ assessment against GPCP is presented in **Fig. 3.3**. Across different seasons, significantly improved skill can be seen over the Southeast Asia (e.g. Myanmar and Thailand), parts of southern China (e.g. the Mei-Yu front), and the Indian subcontinent. The skill assessment for these areas all incorporates the rainfall component after their regional monsoon onset, because the Asian monsoon typically has the earliest onset in the Indochina peninsula in late April to early May, followed by the southern China Mei-Yu in mid-May, and the Indian summer monsoon (ISM) in May to June (Ding and Chan 2005). The exception is the JJA ISM forecast where no signal is detected in MOD, and such a sensor-dependent behavior will be discussed in Section 3.4.4.

A thorough analysis with five reference datasets, RMSE and $r^2$, and statistical tests suggest that among all Asian monsoon regions, central north India (CNI, green boxes in **Fig. 3.9**, 20ºN–28ºN, 74ºE–85ºE) features the most pronounced and robust improvement in the rainfall forecast skill, as the improvement can be consistently seen independent of metrics and rainfall references used. Compared to CNI, signals over other regions are much less robust. CNI receives the most rainfall in the ISM, and an improved area-averaged prediction of ~8% (**Fig. 3.10a**) is quite promising considering its profound agricultural implications and high socio-economic impacts. The Taylor diagram shows that DA can reduce RMSE by up to 30% and increase $r^2$ by up to 15% with our data samples (**Fig. 3.10b**), however, it is important to note that the increase in explained total variance consists of both the interannual and intraseasonal components. A further

decomposition (**Fig. 3.11**) shows the interannual variability, one of the topmost concern for seasonal forecast, is improved in MAM and AMJ albeit sample-limited. In comparison, for longer-lead forecast in MJJ and JJA, there seems to be no improved interannual variability while the bias component is consistently reduced across all seasons (see **Fig. 3.10** for PRMSE). Using information theory, DelSole and Shukla (2010) clarified the model climatological bias and the model skill are interrelated, emphasizing the need to understand the model bias on top of conventional practices of only assessing the skill in seasonal climate forecast studies. To this end, although our limited data samples constrained the quantification on how much snow DA can respectively contribute to the bias and the variance component (especially the interannual variance), the bias reduction still shows promise for the CNI rainfall forecast, but future studies attributing the contribution to each component and understanding their interchanges are needed using more DA data samples.

In the Asian monsoon region, some locations also have degraded skill (e.g. regions at the edge of the Asian monsoon boundary, see blue colors in **Fig. 3.3a**), which may be related to the known issues with Z1416, as well as sampling noises. We note that a perfect improvement is by no means expected because the DA uncertainties and possible atmospheric/sampling noises may still exist that can influence the results. Thus, to more objectively assess the role of DA, we summarize δDA for the Asian monsoon subsystems, in which δDA is computed as the percentage area with improved skill minus that with degraded skill at 95% confidence level (Thomas et al. 2016). A positive (negative) δDA thus suggests improved (degraded) rainfall forecast, while a value of zero suggests no net contribution from DA. For Asian monsoon outlined in **Fig. 3.3a** (areal extent: $11.29\times10^6$ km$^2$), δDA is positive 20%–30% with slightly higher values in GRAMOD (**Fig. 3.3b**). For the ISM (areal extent: $4.3\times10^6$ km$^2$), the strongest signal is seen when δDA exceeds positive 50% during the monsoon transitional period, when the strong EA warming signal starts to dominate in GRAMOD (**Fig. 3.2**). Note that the results for δDA will look unimpressive for regions beyond the ISM and the greater Asian monsoon region, but we only focus on the Asian monsoon areas because of the long-recognized linkage between the Eurasian snow and the monsoonal

rainfall (e.g. Blanford 1884; Senan et al. 2016). Other local and non-local snow–precipitation mechanisms may also exist to help explain the full results but they are not the primary focus of this study, in which we also expect to see less signals if focusing on locally-influenced precipitation due to its greater observational data uncertainties compared to that driven by large-scale circulation changes.

The ISM is the strongest Asian monsoon subcomponent that directly responds to seasonal variations in the land–sea thermal contrast (Kitoh et al. 2013), which may help explain the much more widespread improvement in the ISM than the East Asian Summer Monsoon (EASM; 10°N–30°N, 105°E–125°E). Previous studies showed that the EASM is subject to more complex interplays among different boundary forcings (Dai et al. 2013; Zhou and Zou 2010), thus, the snow DA-forced signal might be potentially disturbed by forcings other than snow DA over Eurasian, which results in the much less widespread improvement in the EASM. In a study investigating the global monsoon's responses in a warming climate, Kitoh et al. (2013) reported that the South Asian Summer Monsoon (SASM) exhibits, among the global monsoon regions, the strongest sensitivity to the land thermal forcings; such conclusions are consistent with our results here, which may ascribe the monsoons' different responses to snow DA to their intrinsic sensitivity to the Eurasian landmass thermal forcing.

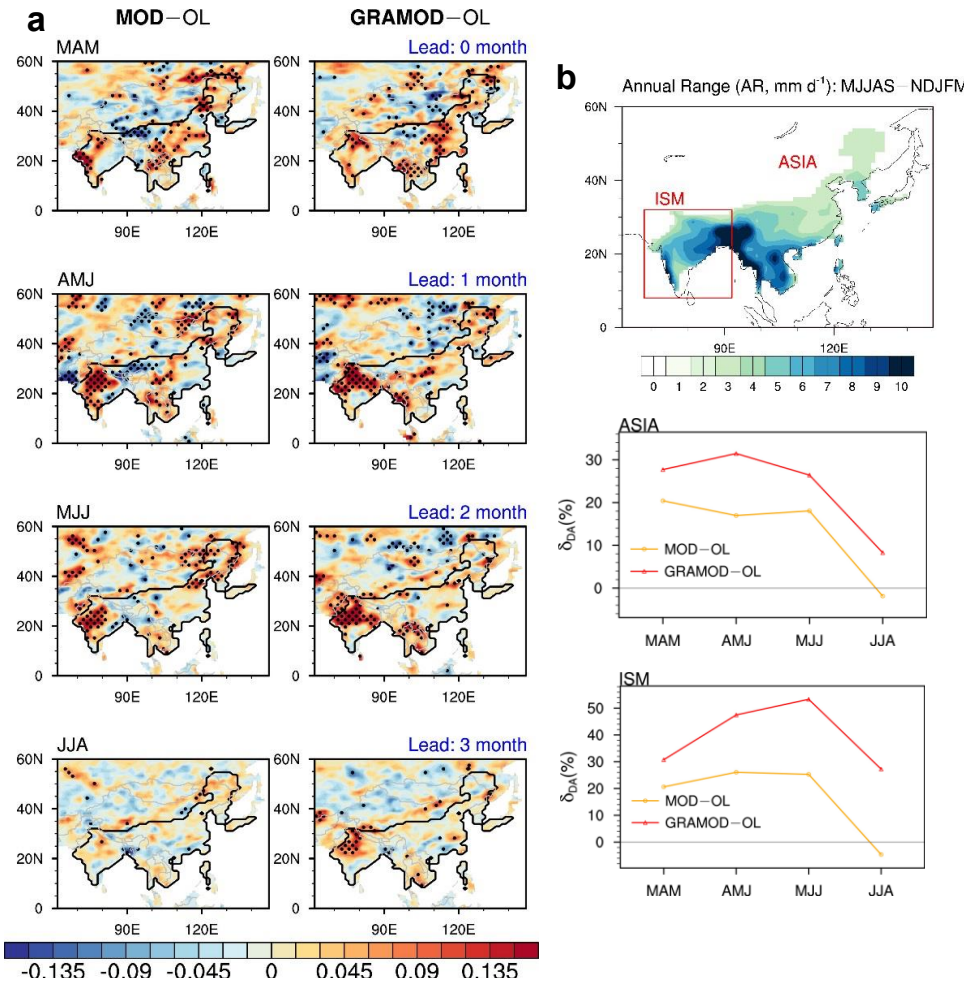


**Figure 3. 3.** Precipitation forecast skill derived from snow DA. (a) shows the skill from MOD (left) and GRAMOD (right) over the Asian monsoon region, which is outlined if the absolute precipitation annual range (MJJAS–NDJFM) exceeds 2.5 mm $d^{-1}$ (Kitoh et al. 2013). (b) shows δDA for the Asian monsoon (ASIA) and the Indian Summer Monsoon (ISM), where δDA is calculated as percentage area with significant improvement minus that with significant degradation (95% confidence level based on bootstrap, see Methods).

### 3.4.4. MODIS and GRACE DA for tailored Asian monsoon forecast

How does the DA uncertainty limit the contribution of DA to the monsoon prediction? To further understand this question, we focus on the Indian subcontinent where the most robust improvement is located, in order to facilitate a discussion on future DA refinements.

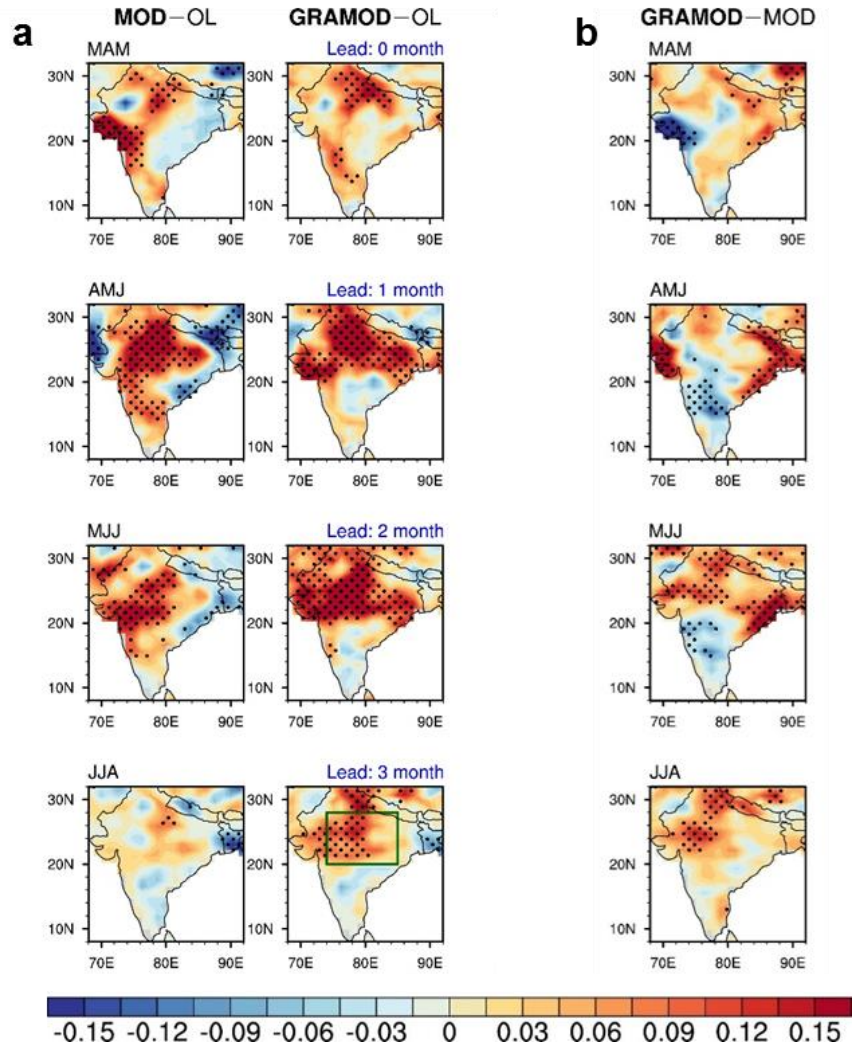


**Figure 3. 4.** Precipitation forecast skill over the Indian subcontinent derived from snow DA. (a) shows the skill from MOD (left) and GRAMOD (right), respectively. (b) shows the exclusive contribution from GRACE DA.

In addition to MOD and GRAMOD (**Fig. 3.4a**), the exclusive contribution from GRACE is isolated by differencing the skill of GRAMOD and MOD (**Fig. 3.4b**). We see that assimilating GRACE data only brings significant improvement to the 2- to 3-month lead forecast in summer (particularly JJA). By contrast, GRACE DA barely offers improvement to the shorter-lead rainfall forecast in spring, and it even slightly offsets the contribution from MODIS DA. Recall that the dominant DA signal comes from the TP in spring (**Fig. 3.2b**), this slightly degraded skill may indicate that the GRACE DA methodology may have room to be improved over the TP. In fact, to

conduct multi-sensor DA in the absence of "true" snow conditions, Zhang and Yang (2016) applied no rules of preference when MODIS and GRACE DA give opposite signs in updating snow. Such an issue occurred over the TP where MODIS DA reduced more snow cover than GRACE DA did (see **Fig. 3.1b**). However, evaluating the DA performance over the TP has proven to be difficult due to the unknown "truth". By focusing on the most robust signal in the coupled land–atmosphere modeling experiments, we show from temperature forecast (Lin et al. 2016) and rainfall forecast (this study) results that the joint GRACE DA methods may be inferior to MODIS DA alone over the TP. The same analyses can be also performed to understand the DA performance over the EA – by focusing on the longer-lead forecast in MJJ and JJA when the EA signal dominates (**Fig. 3.2b**), we see that GRAMOD significantly outperforms both OL and MOD. This highlights the critical role of accurate EA snow initialization constrained by GRACE DA in skillful ISM forecast. Although the GRACE DA method with Z1416 still presents challenges over certain arctic river basins where surface water variations are large but not yet accounted for (Zhang and Yang 2016), further refinement in disaggregating the GRACE TWS signal (e.g. Eicker et al. 2014; Kumar et al. 2016) is expected to improve the skill on top of GRAMOD presented in this study, which implies an overlooked source of the ISM predictability from GRACE DA applied to EA. To address the inconsistent snow updates from MODIS and GRACE over the TP (**Fig. 1c**), however, rule-based DA methods (e.g. Kwon et al. 2016) such as rejecting GRACE while retaining MODIS might be warranted to further advance the skill for tailored ISM forecast.

### 3.4.5. Importance of regional snow DA for the ISM forecast

The previous analyses necessitate a better understanding on the relative importance of regional DA. To address this question, we conduct two more experiments that apply GRAMOD DA only to the TP and EA, respectively (see Section 3.3 and **Fig. 3.12** for details of the two experiments), because they are the two key regions-of-interest for the Asian monsoon studies. Consistent with our above analyses, **Fig. 3.5** shows that the TP snow is indeed more important for the CNI rainfall forecast in spring (MAM and AMJ). In comparison, for the 3-month lead ISM

forecast, the additional skill mainly comes from the EA snow, and the reason is because it only starts to thermally act on the Eurasian landmass and initiate changes in the monsoon circulation in May and June (see **Fig. 3.2b** and **Fig. 3.4** for consistent results). These regional experiments confirm the TP snow as a more important factor for the CNI rainfall forecast during the pre-monsoon season (MAM and AMJ), while emphasizing the importance of the EA snow for the ISM forecast during the monsoon season (JJA). Due to the vicinity and the strong sensible heating at the elevated TP (Wu et al. 2012; Wu et al. 2007), snow over the TP has been widely studied as a significant source of predictability for the Asian monsoon (Turner and Slingo 2011). Our results show that due to its relatively shorter memory (i.e. the forced signal disappears by the end of May, **Fig. 3.2b**), the TP snow and its associated DA can only maintain the influence to the rainfall forecast in the spring. By contrast, the EA snow has a longer memory and its accurate springtime initialization with GRACE data assimilation is more important for the Indian summer monsoon rainfall at 3-month lead. The improvement, however, needs to be further separated in terms of the interannual and intraseasonal variance as well as model biases mentioned in Section 3.4.3, in order to help us better understand the role of DA in informing the monsoon predictability.

The results of this study demonstrate promising additional skill from satellite-constrained snow initializations, which has not been incorporated in current dynamical monsoon predictions. However, it should be emphasized that several potentially overlooked complexities are also clarified in this study that may impede the successful translation of snow DA to the monsoon prediction skill. In addition to the chaotic nature of the monsoon and the DA uncertainty, we find that the regional snow–atmosphere coupling strength is also very important because it determines when and where DA can start to thermally force the atmosphere. This complexity seems to be underappreciated by both the snow DA community and the seasonal forecast community, which constrains our understanding of the monsoon prediction skill attainable from snow DA. Using 7-yr data samples and comprehensive statistical analyses, we also find that the improvement is more robust in central north India of the ISM, the most typical and strongest monsoon subsystems exhibiting large sensitivity to thermal forcings of the Eurasian landmass, compared to other Asian

monsoon regions. These findings should inform future studies using multi-sensor observations to constrain snow initializations, as well as seasonal forecasting systems that attempt to use them for the dynamical monsoon forecast. For example, there are emergent studies using other satellite information to improve the snow estimation, such as those assimilating the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) brightness temperature (Kwon et al. 2017), and jointly assimilating the optical-microwave-gravity observations (Zhao and Yang, 2018, *in revision*). However, to successfully communicate the complementary satellite information to the useful monsoon prediction skill, results from this study suggest that there are more complexities that need to be jointly considered.

Interestingly, we also find some significant improvements in the North African Monsoon (NAF) rainfall forecast (see **Fig. 3.13** for the global monsoon distribution), for which δDA reaches ~35% in MJJ and JJA at 95% confidence level. Some recent monsoon studies reported the possible role of the TP surface heating in influencing its "upstream" NAF climate (Lu et al. 2017), while some viewing the Afro-Asian monsoon as a planetary-scale system potentially responding to the same boundary forcings (Li et al. 2017). Given that the snow DA forcing is applied to the zonally-aligned Eurasian landmass as well as the elevated TP (**Fig. 3.1b**), it is possible that snow DA could also communicate a detectable improvement to the NAF monsoonal prediction. However, it is beyond the scope of this study to clarify the underlying mechanisms that may translate DA to the NAF forecast skill. Future idealized modeling experiments with carefully designed regional DA are warranted in a coordinated multi-model effort, with the intent to shed light on the complex interplay between snow and the global monsoon circulation.

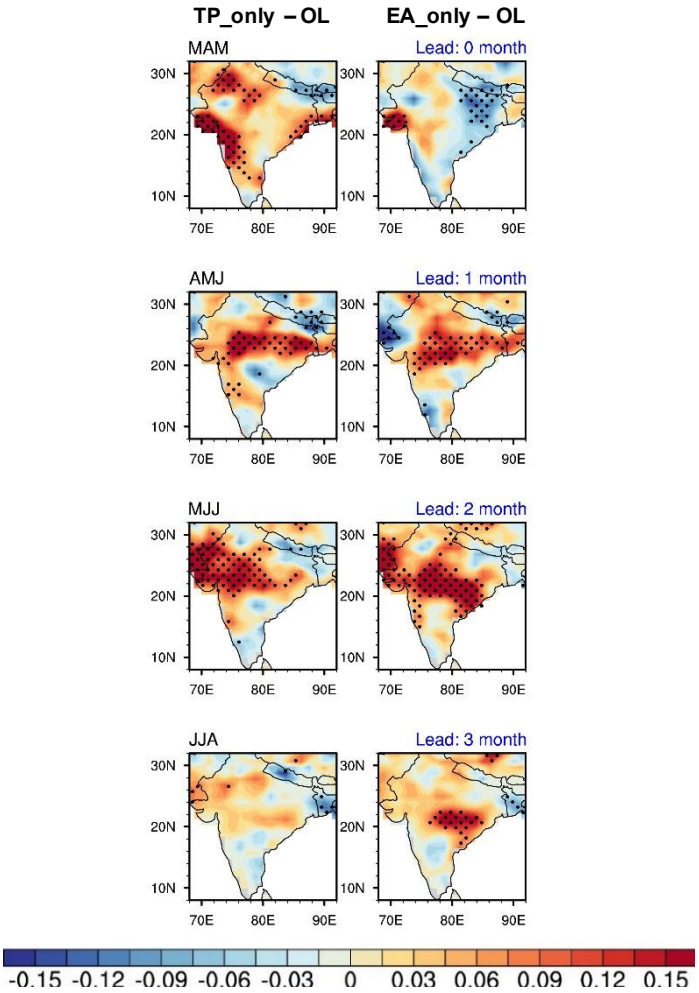


**Figure 3. 5.** Precipitation forecast skill over the Indian subcontinent derived from regional DA. Results from TP only and EA only initializations from the GRACE DA experiment are shown on the left and right, respectively.

## 3.5. CONCLUSIONS

Providing a skillful Asian monsoon seasonal forecast is an unresolved challenge for dynamical climate models, where the predictability has been primarily rooted in the tropical oceans and DA methods to improve the ocean initialization. Here, we show evidence that additional prediction skill can be derived from assimilating multi-sensor satellite data into the land

component to improve springtime snow initialization, which has been an underutilized source of predictability despite numerous efforts focusing on snow DA methodologies.

Due to the thick snowpack and relatively large snow simulation uncertainty, DA updates for the western TP and mid- to high-latitude Eurasia are pronounced – in this study, reduced snow with year-to-year variations are introduced by assimilating MODIS SCF and GRACE TWS. Together, these regions operate on warming up the local atmosphere at different times of the year due to regional snow–atmosphere coupling strength, which then thermally forces the monsoon circulation differently. During early spring, the snow DA signal is dominated by the TP warming; in May and June, the EA warming starts to dominate. The EA warming is much more obviously seen in GRAMOD than in MOD, which then initiates changes to the monsoon circulation and further strengthen, exhibiting sensor-dependent differences when communicating the DA-forced signal to the Asian monsoon forecast. Among the Asian monsoon subsystems, central north India of the Indian summer monsoon exhibits the most prominent and robust improvement signal, which is likely due to its strong sensitivity to the thermal forcing over the Eurasian landmass.

By focusing on the most robust signal seen over the Indian subcontinent and conducting regional DA experiments, we further find that DA over the TP is more important for the pre-monsoon rainfall forecast, and MODIS DA slightly outperforms that with joint GRAMOD DA. In comparison, for the JJA rainfall forecast at 3-month lead, the springtime EA snow initialization plays a dominant role where GRAMOD significantly outperform OL and MOD, highlighting a critical yet underappreciated source of predictability from GRACE DA.

Although this study shows promise for the Asian monsoon seasonal forecast from the snow DA perspective, our results also help clarify some potentially overlooked complexities that limit our understanding of the monsoon forecast skill attainable from DA. In addition to the chaotic nature of the monsoon and the DA uncertainty, we find that the regional snow-atmosphere coupling strength also plays an important role in determining the location and timing for DA to be effective in thermally forcing the overlying atmosphere. The regional monsoon's sensitivity to thermal forcing over the Eurasian landmass seems also important. However, we note that better

mechanistic understanding of the complex interplay between the snow and monsoon circulation is warranted in future idealized experiments using multi-model frameworks. This study identifies the geographic regions where snow DA methods need to be refined for tailored Asian monsoon seasonal forecast. It also clarifies the complexities in obtaining monsoon forecast skill. The results are thus expected to inform future studies seeking for additional sources of monsoon predictability from refined snow DA methodologies.

The statistics presented in this study are mostly constrained by the limited data sample size from the 7-yr snow DA products, which may contain large sampling noises. To address this constraints, future studies using more land DA products and separating anomalous snow years are warranted to improve the robustness of the results presented in this chapter.

### 3.6. SUPPORTING FIGURES

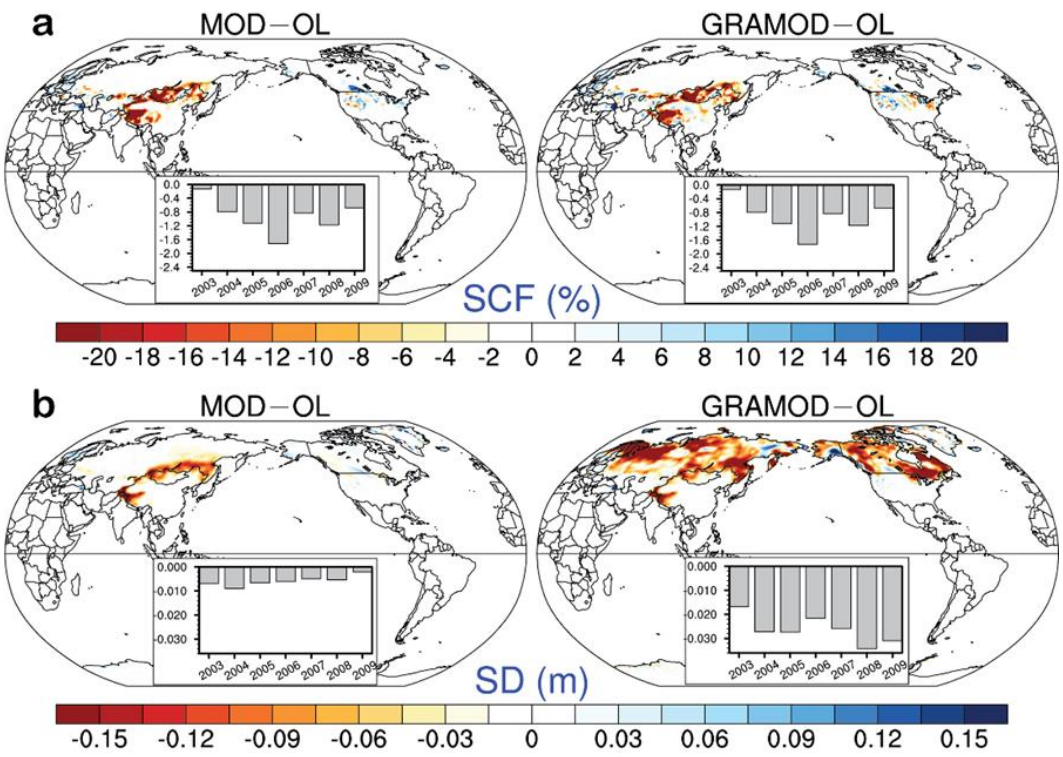


**Figure 3. 6.** DA-introduced changes in initial (a) snow cover fraction (SCF, %) and (b) snow depth (SD, m). Bar charts show the year-to-year variations to indicate the DA effectiveness averaged over the globe.

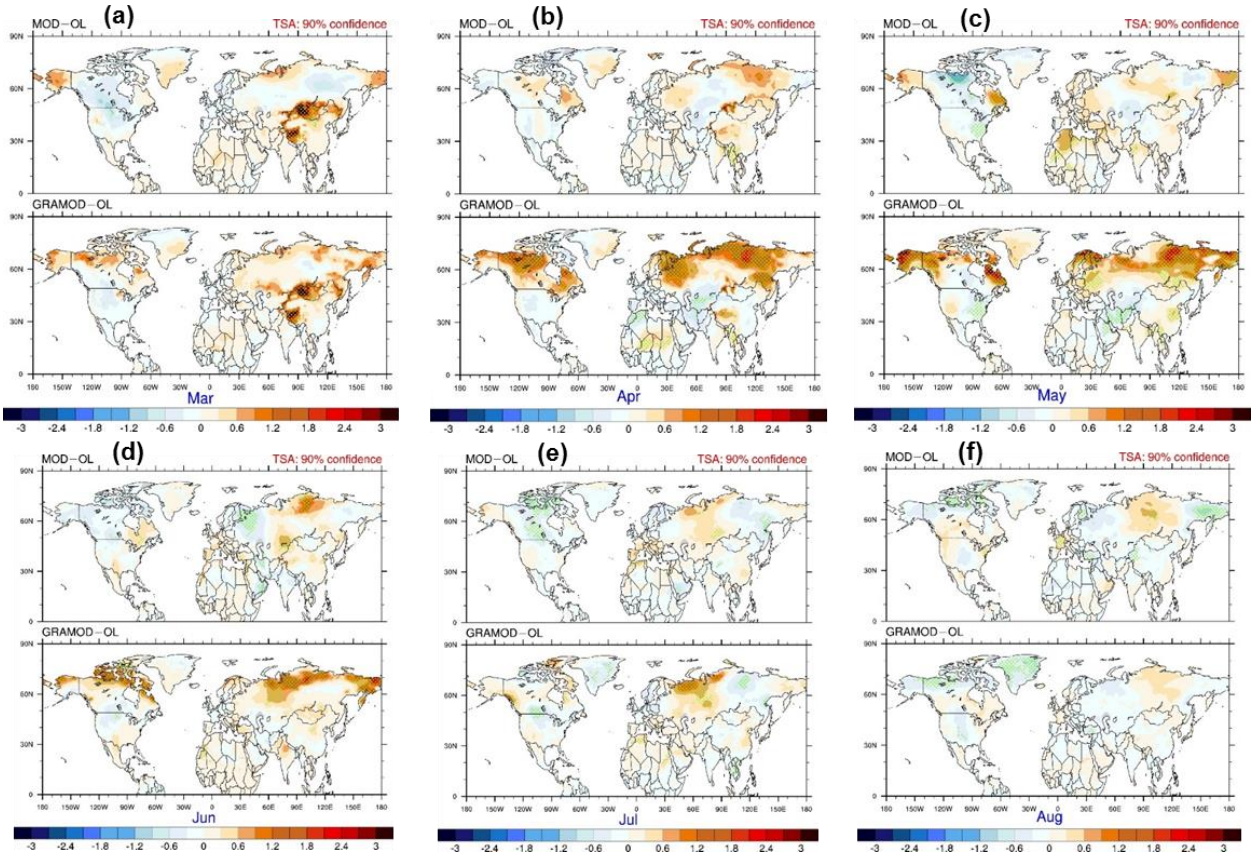


**Figure 3. 7.** Persistence of DA-introduced increase in the surface air temperature. (a)–(f) show the forcing into the forecast period from March to August, respectively; green dots indicate the difference is significant at 90% confidence level.

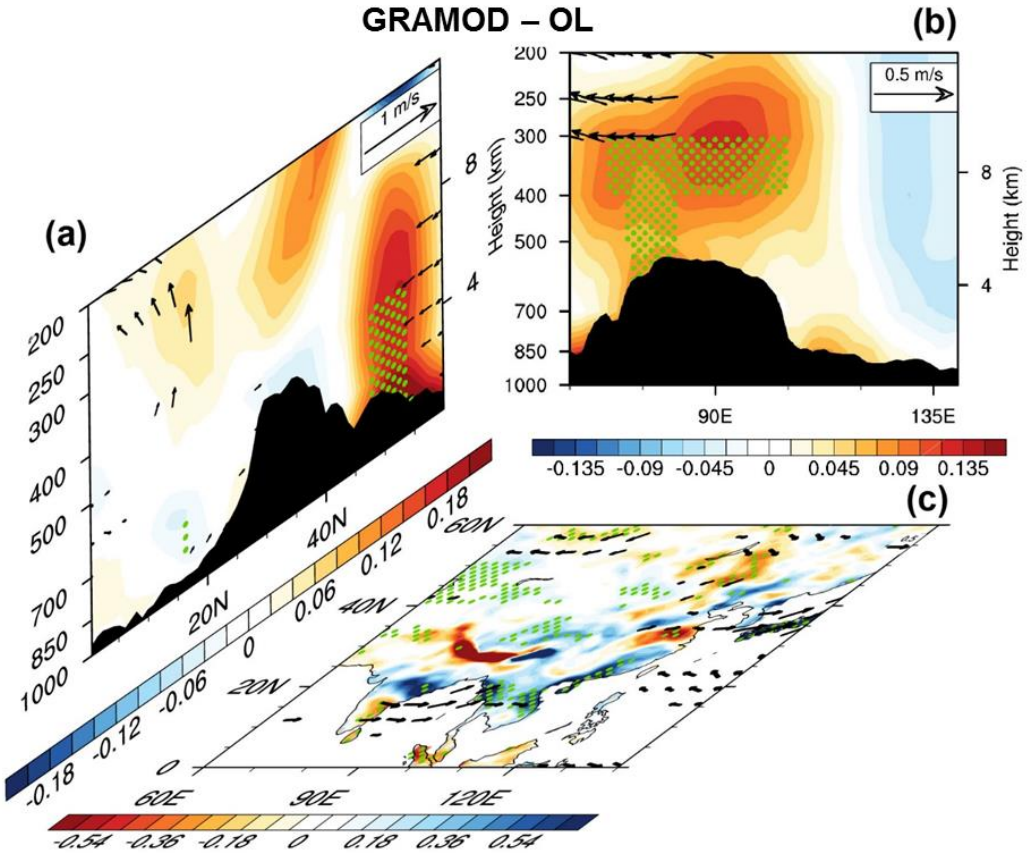


**Figure 3. 8.** GRAMOD-introduced thermodynamic changes during the monsoon season (June–July). (a) (b) show temperature changes (K) in the latitude-height and longitude-height cross sections, respectively, overlaid by meridional-vertical and zonal-vertical wind changes. (c) shows the precipitation changes (mm $d^{-1}$) overlaid by 850 hPa meridional-zonal wind changes. The vertical velocity is multiplied by 400 to be comparable to meridional-zonal wind. The three panels all use the same geographic bounding box. Dots and the vectors drawn indicate the difference is significant at 90% confidence level.

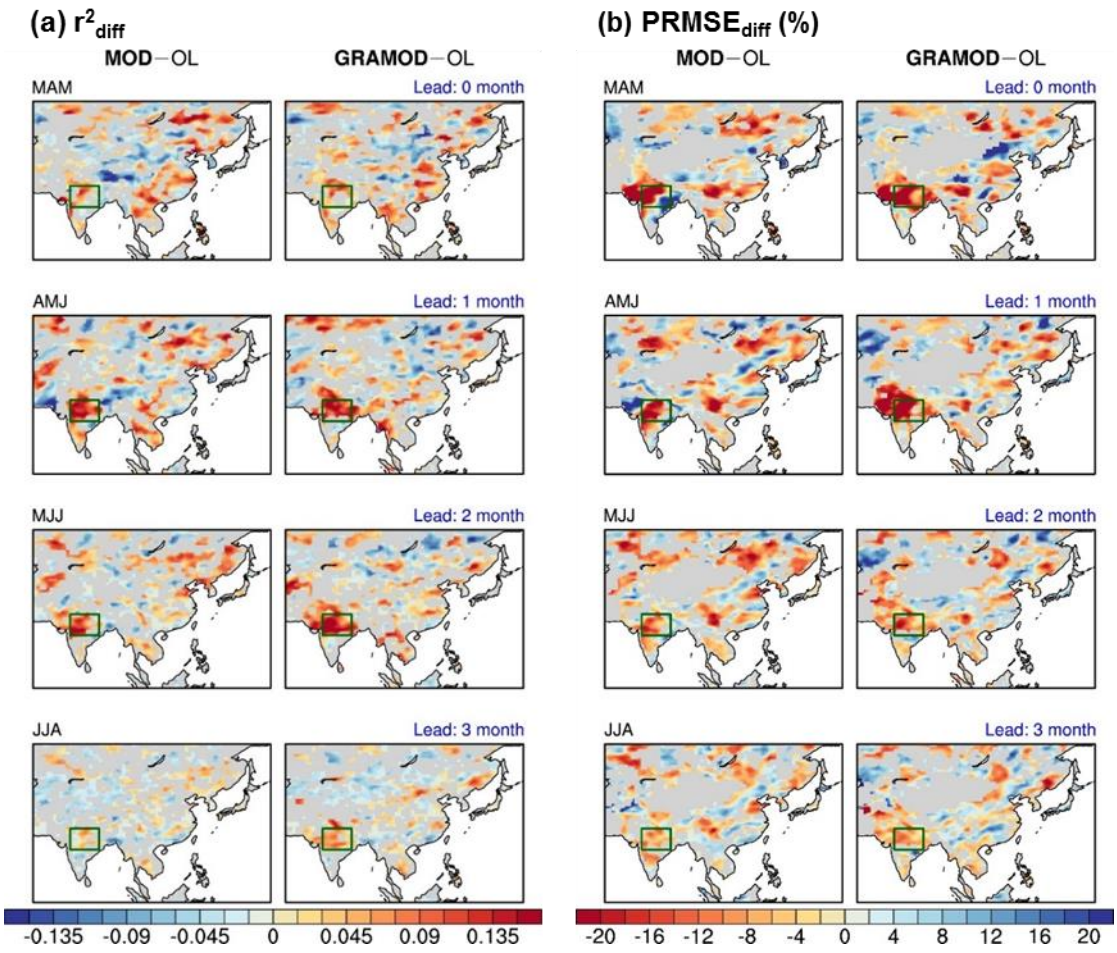


**Figure 3. 9.** Consensus prediction skill against five precipitation reference datasets. (a) shows $r^2_{diff}$ and (b) shows the percentage RMSE ($PRMSE_{diff}$). Both statistics are calculated using 21 samples (see Methods for details). Red colors indicate improvements, and green boxes outline the CNI region.

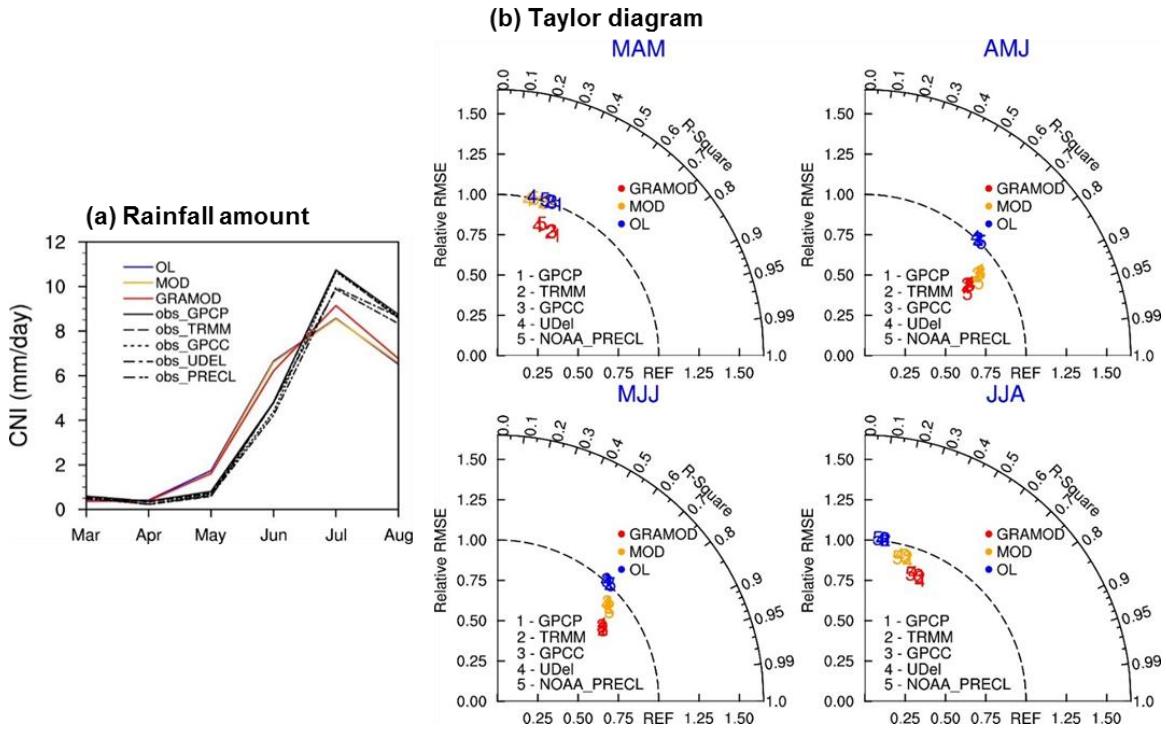


**Figure 3. 10.** Area-averaged CNI rainfall compared against five precipitation references. (a) shows the seasonal cycle of the 7-year average and (b) shows the Taylor diagram with relative RMSE against RMSEOL and $r^2$ based on 21 samples.

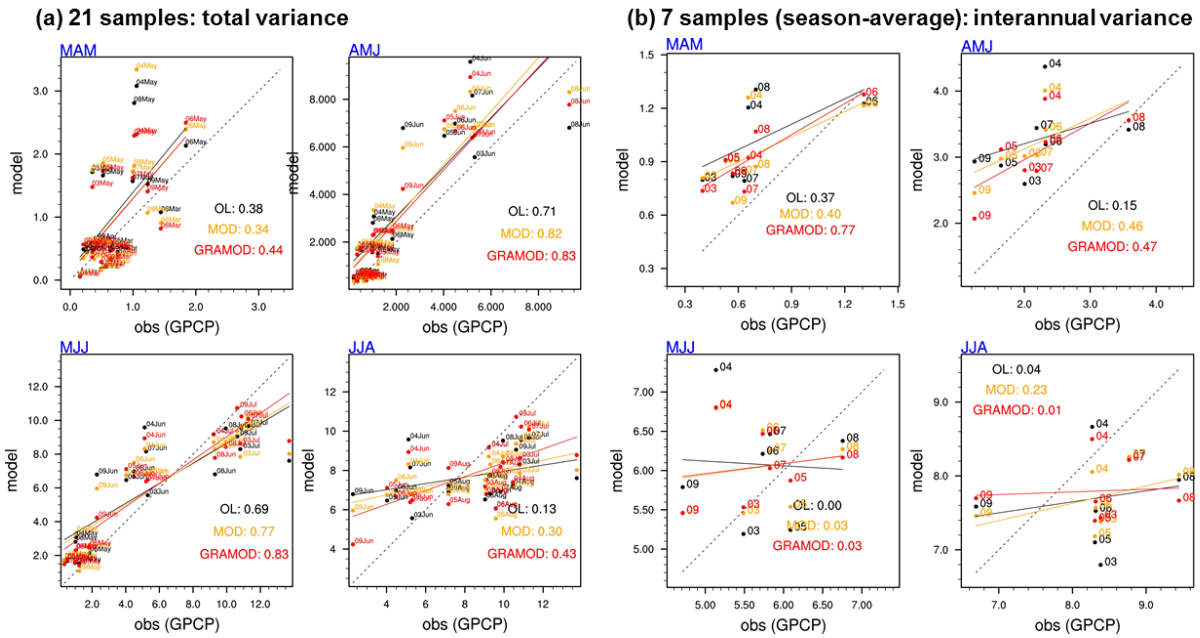


**Figure 3. 11.** Scatter plot for variance analysis using area-averaged CNI rainfall. (a) shows the scatter for 21 samples (7 years×3 months per season) and (b) shows the interannual variability only (7 samples; each dot is a seasonal average). $r^2$ values are shown after the experiment names; texts around the dots show the sample time.

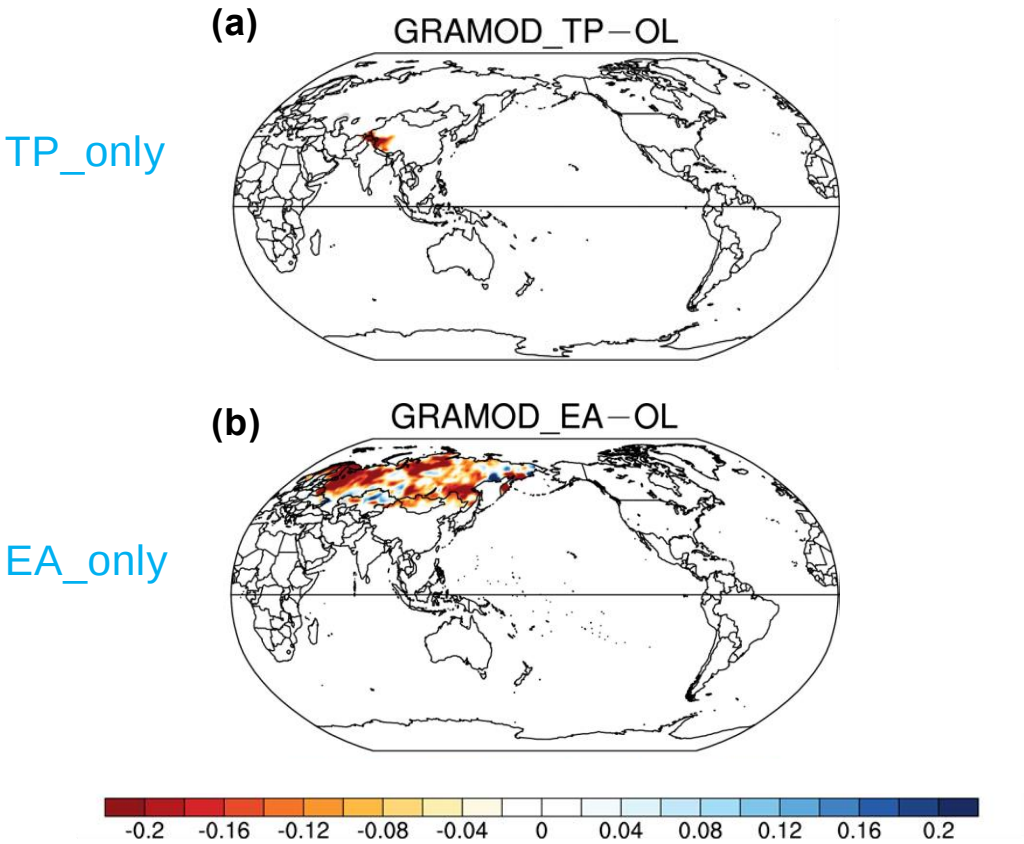


**Figure 3. 12.** Snow depth initialization differences in the regional DA experiments. (a) and (b) show the TP and the EA only experiments, respectively, for the initial snow depth changes (m) introduced by GRAMOD.

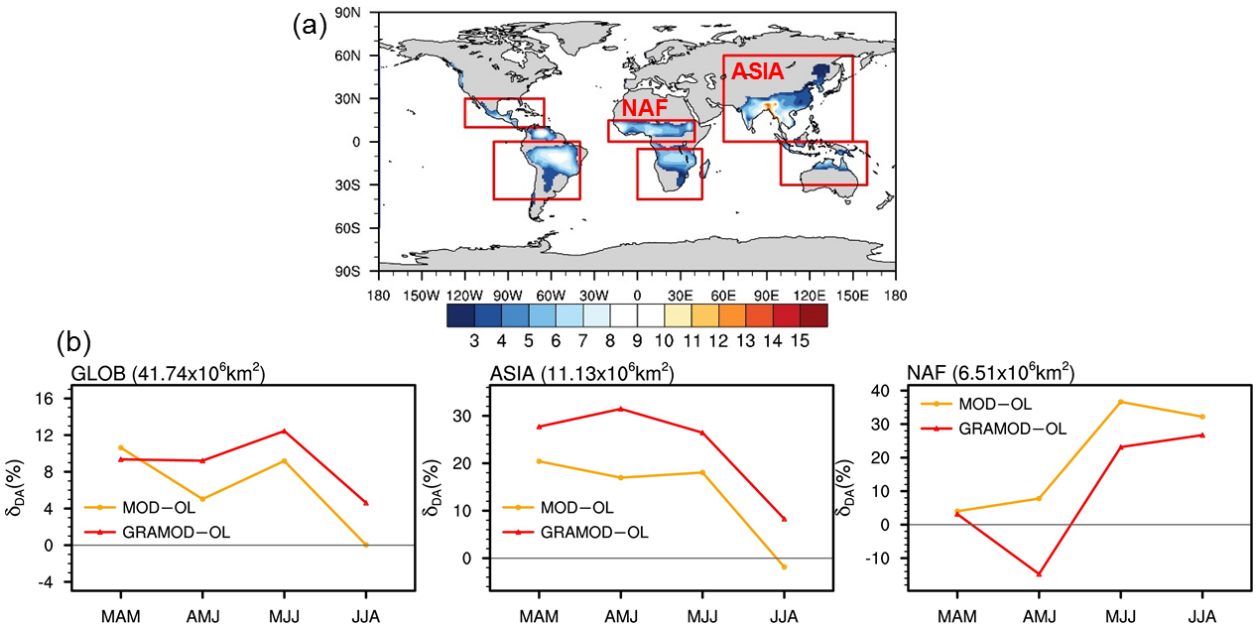


**Figure 3. 13.** Snow DA contribution to other monsoonal areas. (a) delineates the global monsoon regions, where a monsoon domain is defined if the absolute annual precipitation range (difference between MJJAS and NDJFM) exceeds 2.5 mm day$^{-1}$ (Kitoh et al. 2013). (b) shows δDA for global, Asian monsoon, and North African (NAF) monsoon regions, respectively.

## 3.7. ACKNOWLEDGEMENT

This study is funded by the TP project. P.L. acknowledges the University Graduate Continuing Fellowship from UT-Austin. The authors appreciate helpful discussions with Drs. Zhongfeng Xu, Guoxiong Wu, Randal Koster, Yimin Liu, Hui Zheng, and Kai Zhang, Qinjian Jin during the course of this study. The model simulations are performed using the Texas Advanced Computing Center (TACC) supercomputing resources, and all data are available upon request to the corresponding author (liang@jsg.utexas.edu).

# Chapter 4: Implementation of a vector-based river routing scheme in the community WRF-Hydro modeling framework for flood discharge simulation

## 4.1. ABSTRACT

Continental-scale flood discharge modeling requires a high level of efficiency and flexibility. To this end, this study documents the implementation and application of a vector-based river routing model in the community WRF-Hydro modeling framework. By conducting a case study for the inland riverine flood discharge simulation for Hurricane Ike in 2008, the hybrid vector–grid modeling framework's sensitivity to the land grid resolution and the coupling interface is assessed. The evaluation results against discharge observations at 14 gauges show the model is more sensitive to the coupling interface than the grid resolution, and a 1-km land grid with an area-weighted coupling interface exhibits the optimal simulation results. A geographic information system (GIS) based approach is adopted to improve the regional representativeness of the flow travel time estimation. The model's computational efficiency and complexity are compared to a grid-based routing scheme, demonstrating its advantages for large-scale "offline" hydrological applications with GIS-supported features. Trade-offs between the modeling efficiency and complexity are then discussed to inform future flood prediction applications at the large scales. (KEYWORDS: Vector-based river routing; WRF-Hydro; RAPID; Flood discharge simulation; Hurricane Ike; Hurricane Ike)

## 4.2. INTRODUCTION

Vector-based river routing models are receiving increased attention by land surface modeling groups to facilitate large-scale hydrologic predictions at fine resolutions (David et al. 2011; Goodall et al. 2013; Lehner and Grill 2013; Liu and Hodges 2014; Mizukami et al. 2016; Yamazaki et al. 2013). Compared to traditional grid-based routing models, vector-based schemes are conceptually more object-oriented with fewer linkages among objects (Lehner and Grill 2013),

and hence computationally more efficient and generally more accurate in depicting hydrologic features.

Yamazaki et al. (2013) found a 60% increase in computational efficiency when shifting from a grid-based to a vector-based routing approach for a global model CaMa-Flood. In addition to efficiency, vector-based models also demonstrate other advantages. For example, using a vector-based scheme that treats the river network as links and nodes, Liu and Hodges (2014) successfully adapted ideas from microprocessor designs to efficiently solve the fully nonlinear 1-D Saint-Venant equations for routing problems with over $O(10^5)$ elements, which is a traditionally difficult task. Vector-based schemes can provide accurate descriptions of hydrologic features such as river reaches, dams, and gauges at exact geographic locations (Mizukami et al. 2016). In comparison, representations of these features in a grid-based scheme largely depend on the grid resolution, yet increasing the spatial resolution of grid-based models over large domains remains an unachieved priority (Bierkens et al. 2015; Singh et al. 2015; Wood et al. 2011). Using a grid-based scheme, Li et al. (2013) discussed neglecting the within-grid routing may benefit from a vectorized modeling environment. Efficient large-scale hydrologic prediction at fine resolutions could potentially benefit from a hybrid modeling framework that combines both vector-based and grid-based modeling units. However, coupling vector-based routing with traditional grid-based LSMs remains a challenging task, where the former has been mostly practiced with river channel design and floodplain analysis by civil engineers and the geographic information system (GIS) modeling groups (Clark et al. 2015; Li et al. 2013; Mizukami et al. 2016).

Kauffeldt et al. (2016) summarized, after reviewing 24 hydrological models, several key criteria that make a modeling system a suitable candidate in large-scale hydrologic applications in an operational setting. Those similarly motivated the model coupling work in this study include availability of model code, existing user community, flexibility in resolution, as well as flexibility in grid structures and modeling units. In light of these postulated criteria, we implemented the Routing Application for Parallel computatIon of Discharge model (RAPID) (David et al. 2011) into the community hydrological extension of the Weather Research and Forecasting model

(WRF-Hydro) (Gochis et al. 2015) currently in a "one-way" manner, meaning that there is no feedback from the vector-based routing model to the LSM. These two open-source models were developed for different user communities. WRF-Hydro is an architectural framework for broad-scope hydrometeorological studies (Arnault et al. 2016; Gochis et al. 2015; Yucel et al. 2015), while RAPID is mainly practiced with GIS-based water resource management and flood-related applications (Follum et al. 2017; Lin et al. 2015; Tavakoly et al. 2017). The advantages in combining these models were first demonstrated by the National Flood Interoperability Experiment (NFIE) for a continental-scale application (Maidment 2017), where the 15-hr lead streamflow forecasting for 2.7 million river reaches across the continental United States (US) were completed within ~10 min using 16 computing cores at the Texas Advanced Computing Center (TACC) Stampede cluster (Lin et al. 2018; Maidment 2017; Salas et al. 2018). Although the "loosely" coupled modeling strategies (i.e., the outputs of a model are generated, processed, and then used as inputs to another model) demonstrated the computational efficiency, the need for a "tightly" coupled WRF-Hydro-RAPID software package lies in its flexibility in data structure to allow for researchers with interdisciplinary backgrounds to conduct flood prediction studies. In addition, such a modeling system may also make it straightforward to address uncertainties originating from weather inputs, land surface parameterizations, and river routing schemes.

The first objective of this study, therefore, is to document the coupling, implementation, and configuration of the WRF-Hydro-RAPID modeling system. We then assess the system's performance in terms of simulating flood discharge and the sensitivity to land grid resolutions and grid-to-vector coupling strategies using Hurricane Ike as a case study. Radar-based rainfall data are used as the rainfall input and a GIS-based approach is adopted to estimate the flow travel time for the regional modeling simulation. Finally, the trade-offs between model efficiency and complexity are discussed by comparing RAPID with the grid-based routing option in WRF-Hydro, in order to inform future model development and application activities.

## 4.3. MODEL DESCRIPTIONS AND COUPLING

### 4.3.1. The community WRF-Hydro modeling framework

WRF-Hydro is the hydrological extension of the WRF model, a widely-used numerical weather prediction (NWP) tool for weather/climate research and operations. Developed by the National Center for Atmospheric Research (NCAR), WRF-Hydro provides a model coupling architecture to link multi-scale hydrometeorological processes (Gochis et al. 2015), which can be broadly categorized as (1) the atmospheric models and meteorological input data stubs, (2) LSMs for calculating land states and fluxes in 1-D vertical columns, and (3) routing modules for representing horizontal water movement (**Fig. 4.1**). Each of these modular layers can be set active or inactive to allow for stand-alone or full-feature modeling experiments by modifying the "hydro.namelist". The supported full-feature modeling (grid-based) currently includes switch-activated modules for surface overland flow, saturated subsurface flow, base flow in conceptual groundwater buckets, and channel flow processes. Several state-of-the-art LSMs (e.g. Noah and Noah-MP) are supported to benefit multi-model predictions. To this end, WRF-Hydro is not a hydrological model but instead is an extensible modeling framework to serve both "coupled" (Fig. 4.1a, brown arrows) and "offline" (Fig. 4.1b, green arrows) hydrometeorological applications.

Utilizing these modeling capabilities currently supported by WRF-Hydro, previous studies have investigated the "two-way" interactions between soil moisture redistribution and hillslope hydrology–atmosphere feedbacks in West Africa and the central Mediterranean (Arnault et al. 2016; Senatore et al. 2015), while some focused on "offline" hydrological applications to study the "one-way" hydrologic response to the weather forcings (Silver et al. 2017; Yucel et al. 2015).

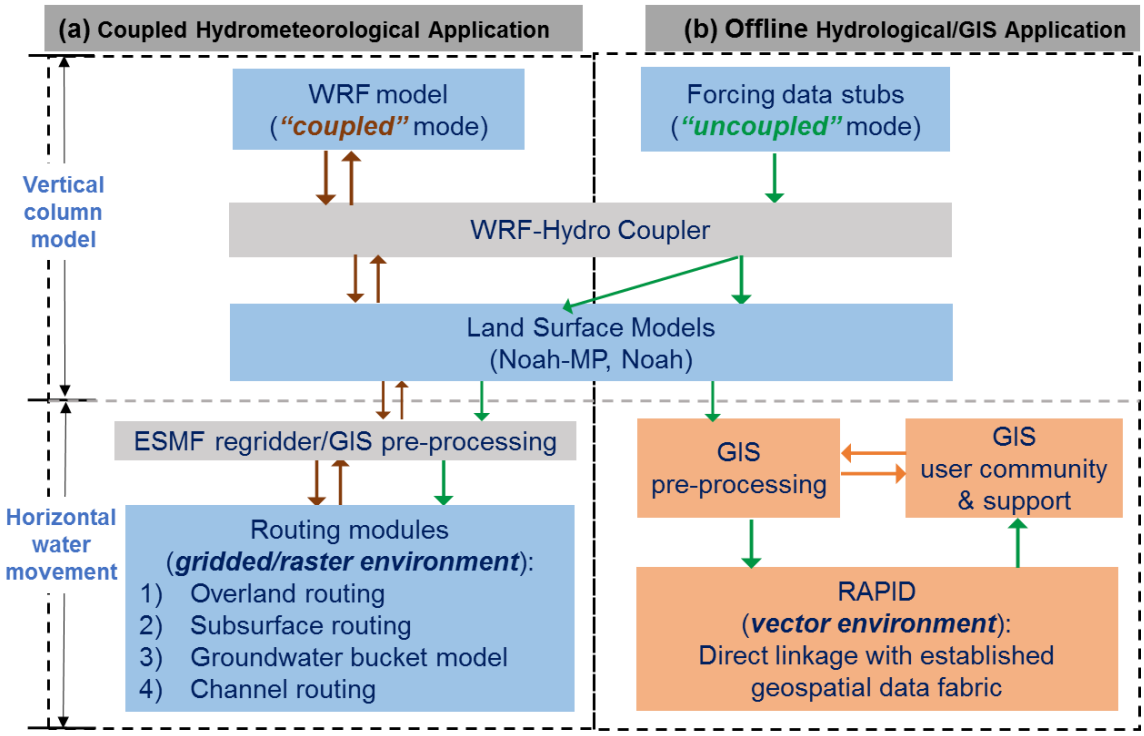


**Figure 4. 1.** Schematic diagram of the WRF-Hydro architecture and the newly implemented vector-based routing scheme (orange). (a) has "two-way" communications between the weather models and the hydrological components (brown arrows); (b) is "one-way" or "offline" hydrological application that uses weather forcings to drive the hydrological components (green arrows).

The grid-based routing in WRF-Hydro separates the explicit representation of terrain and channel routing processes (Gochis et al. 2015), which is similar to Li et al. (2013). Before water reaches a predefined channel cell, switch-activated overland, subsurface, and baseflow bucket modules are used to route the flow, which allows for updating surface head along the pathway (dashed black arrows in **Fig. 4.2a**); once the water reaches the predefined network, the channel routing module takes over to route water down. The routing performance of such a grid-based scheme is thus sensitive to both the land and the channel grid resolutions (**Fig. 4.2a**). For example, given a fixed land grid resolution, a coarser channel resolution (larger grid spacing) can resolve less tributaries of lower Strahler orders, resulting in lower tributary density in one LSM grid cell. In this case, terrain routing becomes more important due to a lack of channel pathways to route the water compared to the case with finer channel resolution. In this study, our focus is on

augmenting the "offline" hydrological modeling capability of WRF-Hydro with a vector-based routing scheme, which has no requirement for a grid-based channel mesh.

### 4.3.2. The RAPID river network routing model

RAPID is a river network routing model that can directly perform streamflow computations on the vector-based geospatial data fabrics, such as a continental-scale dataset from NHDPlus (David et al. 2013; David et al. 2011; Follum et al. 2017; Lin et al. 2015; Tavakoly et al. 2017), and a global-scale dataset from HydroSHEDS (David et al. 2015; Wood et al. 2011). These two datasets both provide "blue-line" hydrography for use in modeling and efficient hazard communications. Different from other models that also originated from the GIS community such as the Soil and Water Assessment Tool (SWAT), RAPID by design may be more easily integrated with traditional grid-based LSMs in a "tight" coupling strategy, because of its modular structure written in FORTRAN and its ability to handle large-scale routing problems. The RAPID version we implement in this study supports both the traditional and matrix-based forms of the Muskingum method for the reach-by-reach channel routing. The vector-matrix formulation of RAPID stores the river network as a matrix of links with certain degree of connectivity, then the Muskingum equations (in matrix form) are solved with the finite-difference form of the continuity equation with a linear system solver, which can avoid spatial iterations and support parallel speedup (David et al. 2011). Since both the United States National Water Model (NWM) and this study use the medium-resolution NHDPlus hydrography dataset (1:100,000) for neighborhood-scale hydrologic applications, this dataset is described with more details below.

NHDPlus is an integrated suite of geospatial dataset developed by the United States Environmental Protection Agency (EPA) and the United States Geological Survey (USGS) (http://www.horizon-systems.com/nhdplus/). It includes the river flowlines extracted from the 30-m National Elevation Dataset (NED), where significant field survey corrections have been applied to improve their accuracy over the past two decades (Moore and Dewald 2016). Taking the Texas Hydrologic Region 12 as an example, the medium-resolution NHDPlus contains 68,143 river

reaches. In comparison, it would require more than 1 million grid cells in a grid-based framework to delineate these channels at comparable accuracy, which suggests a reduction by two orders of magnitude in the number of modeling units using a vector-based routing model.

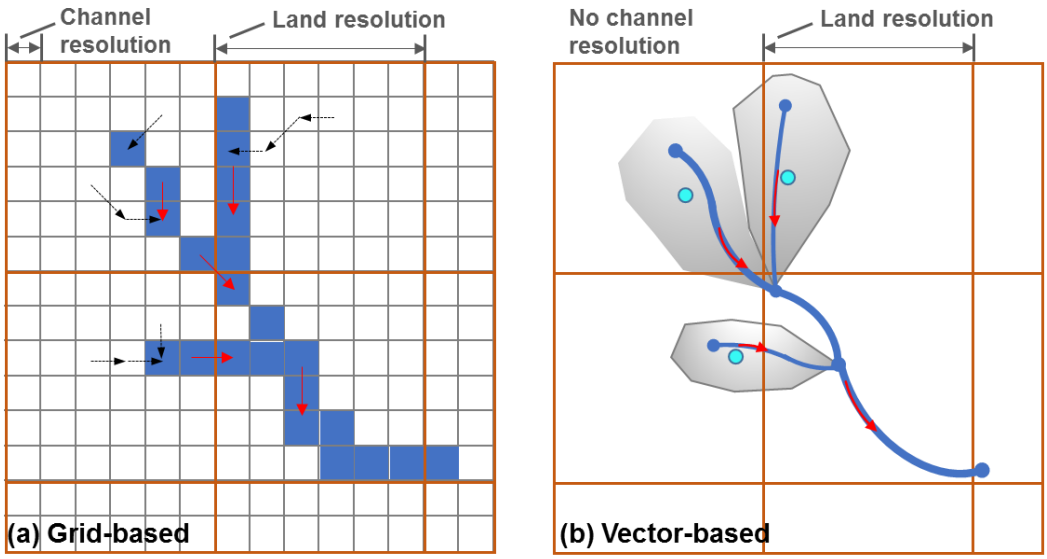


**Figure 4. 2.** A conceptual illustration of the difference between (a) grid-based and (b) vector-based routing schemes. The grid-based routing has 23 channel cells as the basic modeling units in (a) while having 5 river reaches as the basic modeling units in (b). Red arrows show the channel flow directions; black arrows show the terrain routing before water reaches the pre-defined channels.

**Fig. 4.2** illustrates the conceptual modeling strategy difference between the grid-based and vector-based routing schemes. While the latter performs the streamflow simulation on the vector-based reaches with more accurate hydrologic delineations and less modeling units, it may also be subject to less spatial fidelity than a grid-based scheme that can be implemented using sufficiently high channel resolutions. For example, **Fig. 4.2a** shows a grid-based scheme with high channel resolutions (assuming 30-m). In this case, the river network is delineated at about the same accuracy as in **Fig. 4.2b**, but each channel modeling unit in **Fig. 4.2a** has an explicit resolution of 30 m, while each modeling unit in **Fig. 4.2b** has a resolution of ~150 m (based on the average reach length). This suggests a possible trade-off between computational efficiency and spatial

fidelity between grid- and vector-based routing schemes that may influence the time stepping needed for model stability (see Section 4.6). Nevertheless, before grid-based models can be implemented at fine resolutions over large spatial domains, a task yet to be accomplished (Singh et al. 2015), vector-based models still demonstrate potential advantages and flexibilities that warrant a hybrid modeling strategy.

### 4.3.3. Grid-to-vector coupling interface

To implement RAPID as an alternative routing option, a FORTRAN module named "the RAPID coupler" was added to the WRF-Hydro source code, which determines the location where each vector-based river reach can get lateral inflows from LSM-simulated total runoff (we refer tit as the "grid-to-vector coupling interface"). We also augment a simpler catchment centroid-based coupling interface used by David et al. (2013) (**Figs. 4.3(b)(d)**), with an area-weighted coupling shown in **Figs. 4.3(c)(e)**. In the former approach, the LSM grid cell where the catchment centroid is located (hereafter "centroid cell") is treated as the location for a river reach to get lateral inflows. For example, for the river reach in red (**Fig. 4.3(a)**), this implementation could lead to reasonable representations of contributing runoff in **Fig. 4.3(b)**, while problems could occur with an increase in LSM resolution (**Fig. 4.3(d)**). For reaches with consistently small catchment areas (e.g. southern parts in **Fig. 4.3(a)**), however, such variations in the interface seems to be less important for calculating the lateral inflows because the LSM grid resolution is coarse regardless.

If a catchment centroid-based interface option (**Figs. 4.3(b)(d))** is chosen (implemented as an option in the "rapid.namelist"), the contributing lateral inflows ($\boldsymbol{I}_i$, $m^3$) is calculated as:

$$\boldsymbol{I}_i = runoff(nx, ny) \times area \times 1000.0 \qquad \text{Eqn. (4.1)}$$

where (nx,ny) is the indices of the centroid cell relative to the LSM grids; $area$ is catchment area ($km^2$); $runoff$ is LSM-simulated total runoff (mm, surface + subsurface). If the area-weighted coupling interface (**Figs. 4.3(c)(e)**) is chosen, $\boldsymbol{I}_i$ is computed (Eqn. 4.2, $m^3$) by adding up the runoff from all intersected LSM grid cell j ($runoff_j$, mm), multiplied by the intersected area ($area_{intercept,j}$, $km^2$) and ratio of the intersected area in the total catchment area ($weight_j$); n

denotes total number of intersecting cells. The geographic correspondence (nx, ny, j) and the catchment areas are obtained through a series of GIS pre-processing procedures to generate several pre-defined csv files (see Section 4.4.2.2). The added module then reads the information when initializing RAPID (only once), after which the LSM-simulated runoff is dynamically passed to the NHDPlus river reaches at every model time step.

$$\boldsymbol{I}_i = \sum_{j=1}^{j=n} runoff_j \times area_{intercept,j} \times weight_j \times 1000.0 \qquad \text{Eqn. (4.2)}$$

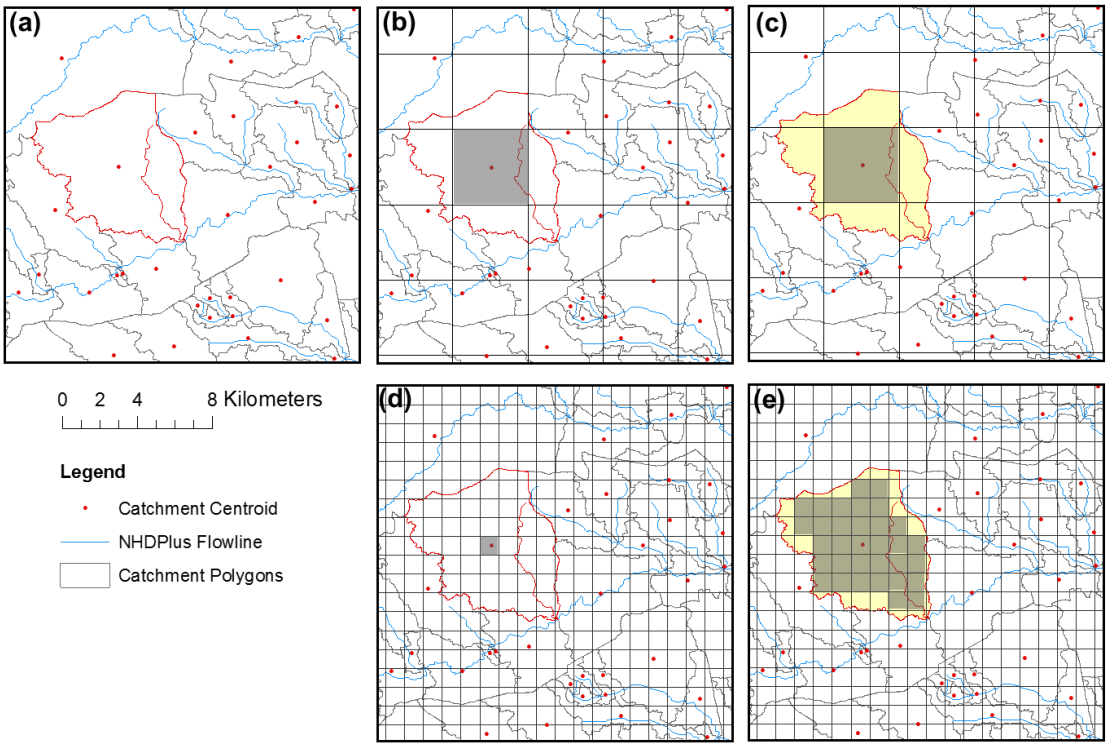


**Figure 4. 3.** An example of the WRF-Hydro and RAPID coupling interface based on the NHDPlus geospatial data framework (shown in ArcGIS). (a) highlights the example river reach and catchment in red; (b) and (c) show the interface with a 4-km land grid resolution, using a centroid-based approach (b) and an area-weighted approach (c). (d) and (e) are similar to (b) and (c), except for using a 1-km land grid resolution. The shaded area shows the land surface model cells where the example river reach extracts the lateral runoff from.

To assess the model performance and its regional simulation sensitivity to the LSM grid resolution and the coupling approaches in such a hybrid framework, we then present a real-world application to systematical assessed these aspects. Spatially-varied estimates of flow travel time is also evaluated to improve the model's performance. Next, the model setup, implementation, and performance for the real-world application are documented to guide future modeling activities.

## 4.4. A CASE STUDY FOR HURRICANE IKE FLOOD DISCHARGE SIMULATION

### 4.4.1. Hydrometeorological analysis

We perform a case study to simulate the inland riverine flood discharges resulted from Hurricane Ike that hit the Texas Gulf Coast in 2008. According to the National Oceanographic and Atmospheric Administration (NOAA), Ike was the third costliest mainland United States tropical cyclones from 1900 to 2013 that was blamed for 21 direct and 64 indirect fatalities and $30 billion (2008, USD) (NOAA, 2009; http://www.aoml.noaa.gov/hrd/tcfaq/costliesttable.html). Hurricane Ike made its landfall near Houston, Texas, at 0700 UTC on September 13 at a maximum wind speed of ~48 m $s^{-1}$ (Zhao and Xue 2009). It intercepted with a cold front as it moved northeastward and crossed the Texas-Louisiana border on September 14, after which Ike gradually decayed going into the Midwest United States and the eastern Great Lakes regions two days after its landfall.

The National Center for Environmental Prediction (NCEP) StageIV quantitative precipitation estimation (QPE) product and the USGS gauge observations are used to first provide a hydrometeorological analysis of the event before setting up the model domain. **Fig. 4.4(a)** shows the rainfall intensity spatial pattern (mm $hr^{-1}$) at 1200 UTC on September 13 and 14, respectively while **Fig. 4.4(b)** shows the 24-hr accumulated rainfall amount (mm) that observed widespread heavy (4–16 mm $hr^{-1}$) to very heavy rainfall (16–50 mm $hr^{-1}$) on September 13. On September 14, the interception with the cold front resulted in geographically-confined heavy rainfalls, but the localized rainfall intensity was much more severe (>50 mm $hr^{-1}$) than the first day. As the most impacted region, the San Jacinto River Basin (SJRB, outlined in dark green) received more than

300 mm of 24-hr accumulated rainfall on the first day (**Fig. 4.4b**), and the basin-averaged 24-hr accumulated rainfall being 172 mm on September 13 and 63 mm on September 14 (**Fig. 4.4c**).

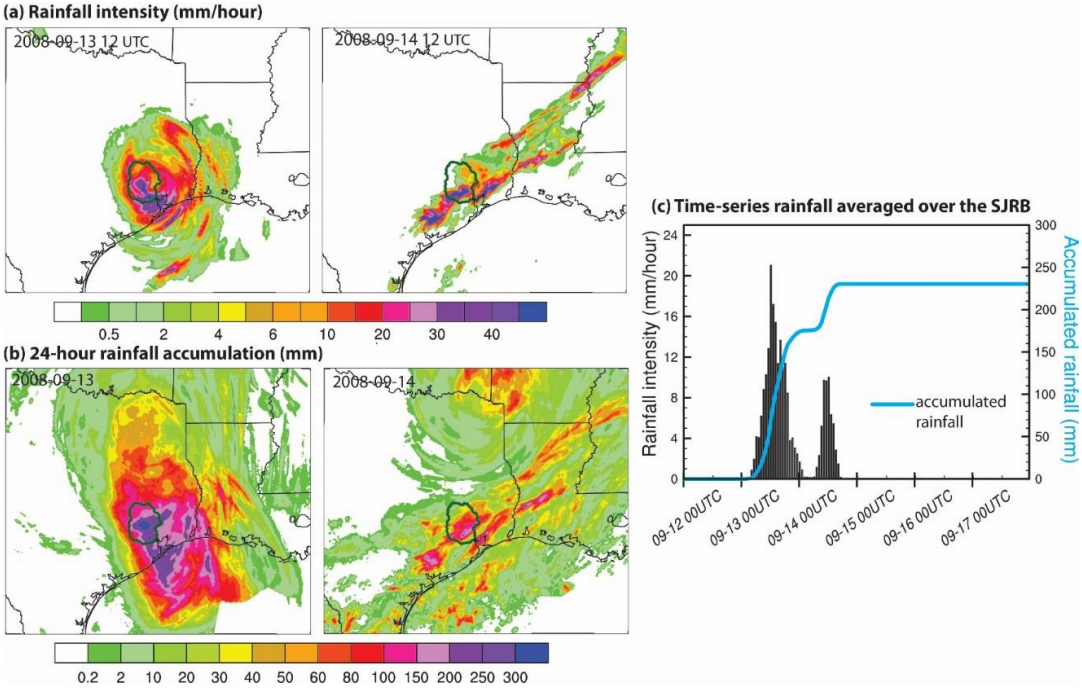


**Figure 4. 4.** The rainfall spatial pattern as shown by the 4-km StageIV product: (a) shows the rainfall intensity (mm $hr^{-1}$) at 12 UTC of September 13 and 14; (b) shows the 24-hour accumulated rainfall (mm) on September 13 and 14, respectively, with cycle ending at 00 UTC. (c) shows the rainfall time series averaged over the San Jacinto River Basin (SJRB, dark green outline).

We conduct a flood frequency analysis (FFA) at all available USGS gauges in the Texas Hydrologic Region (hereafter Region 12) to understand the regional hydrologic response resulting from the event's rainfall spatiotemporal structures, land surface properties, and drainage network structures (Wright et al. 2014). The FFA is conducted using the USGS PeakFQ v7.1 software (Veilleux et al. 2014) that fits the historical annual maximum discharges to a log Pearson Type III distribution using the Bulletin 17B regional skew coefficients. Region 12 has more than 700 USGS gauges, and 237 gauges with more than 25-year historical annual maximum discharges for reliable FFA are used for the analysis (**Fig. 4.5**). In particular, the SJRB has 18 such gauges that have sufficient data, and are active and not subject to known water diversions.

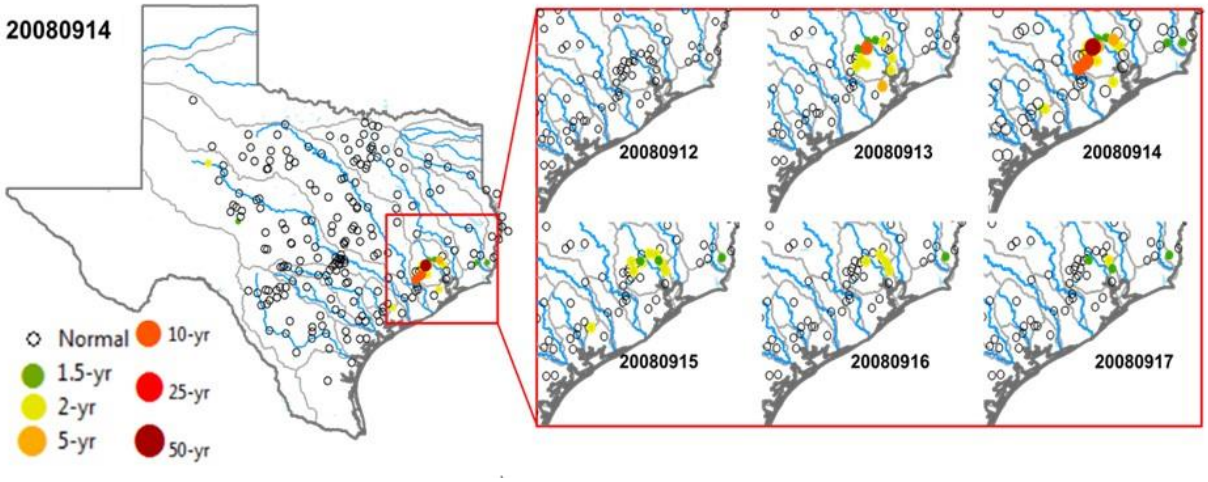


**Figure 4. 5.** Flood frequency analysis at 237 USGS gauges in the Texas Hydrologic Region 12. Gauges with daily maximum streamflow exceeding the 1.5-yr, 2-yr, 5-yr, 10-yr, 25-yr, and 50-yr return period flood discharges are shown in colored dots. A gauge is shown to be under normal conditions if its maximum instantaneous discharge does not exceed that of a 1.5-yr flood (denoted using black circles).

Before the hurricane made landfall on September 12, all locations in the SJRB were under normal flow conditions, meaning that their maximum instantaneous river discharge of the day was smaller than that of a 1.5-yr flood level. On September 13, ten gauges observed 1.5- to 2-yr floods, one gauge observed a 5-yr flood, and one gauge observed a 10-yr flood. The most widespread and severe flooding happened on September 14, where several gauges observed 10-yr to 50-yr floods. Since September 15, the flow conditions at most gauges recessed below the 2-yr flood level, but parts of the Brazos, Trinity, and Lavaca River Basins continued to observe 10-yr floods. Note that the basin-averaged 24-hr accumulated rainfall on September 13 was twice the magnitude of that on September 14, but the most severe 50-yr flood only occurred on the second day. Such a hydrological response may be explained by the localized high rainfall intensity (>50 mm $hr^{-1}$) from the cold front interception with Ike on September 14 (**Fig. 4.4a**, right). In may also be ascribed to the buffering effect of the soil moisture storage in the northwestern part of SJRB, which may have allowed portions of the rainfall to infiltrate (0.2 $m^3$ $m^{-3}$) rather than running off before the soil became nearly saturated on the second day (~0.35 $m^3$ $m^{-3}$; based on the model outputs not shown).

| | Name | Gauge ID | River ID | Location (°N, °W) | Drainage ($km^2$) | Return Period | Peak Date |
|---|---|---|---|---|---|---|---|
| **1** | Willow Creek near Tomball | 08068325 | 1508279 | 30.11, 95.55 | 106.19 | 50-yr | 09/14 |
| **2** | Bear Creek near Barker | 08072730 | 1440237 | 29.83, 95.69 | 55.68 | 10-yr | 09/14 |
| **3** | Langham Creek W Little York Road | 08072760 | 1440211 | 29.87, 95.65 | 63.71 | 10-yr | 09/14 |
| **4** | Buffalo Bayou near Katy | 08072300 | 1440311 | 29.74, 95.81 | 163.95 | 10-yr | 09/14 |
| **5** | Panther Br near Spring | 08068450 | 1508045 | 30.13, 95.48 | 89.35 | 10-yr | 09/13 |
| **6** | Peach Creek at Splendora | 08071000 | 1520237 | 30.23, 95.17 | 303.03 | 5-yr | 09/14 |
| **7** | Spring Creek near Spring | 08068500 | 1508043 | 30.11, 95.44 | 1059.31 | 5-yr | 09/14 |
| **8** | Luce Bayou above lake Houston | 08071280 | 1520091 | 30.11, 95.06 | 564.62 | 2-yr | 09/14 |
| **9** | San Jacinto River near New Caney | 08070200 | 1520083 | 30.15, 95.13 | 1004.92 | 2-yr | 09/14 |
| **10** | Spring Creek near Tomball | 08068275 | 1508047 | 30.12, 95.64 | 481.74 | 2-yr | 09/14 |
| **11** | Little Cypress Creek near Cypress | 08068780 | 1508121 | 30.02, 95.70 | 106.19 | 2-yr | 09/14 |
| **12** | Whiteoak Bayou at Alabonson Road | 08074020 | 1438725 | 29.87, 95.48 | 89.35 | 2-yr | 09/13 |
| **13** | Caney Creek near Splendora | 08070500 | 1520249 | 30.26, 95.30 | 271.95 | 2-yr | 09/15 |
| **14** | San Jacinto River near Canroe | 08068000 | 1468280 | 30.24, 95.46 | 2144.51 | 1.5-yr | 09/13 |

**Table 4. 1.** Summary of the 14 flooded gauges in SJRB.

Based on the regional FFA analysis, 14 gauges experiencing inland riverine floods are separated from those with urban ponding and coastal storm surges, i.e., flood types that are not expected to be captured by the current modeling framework. Therefore, these 14 gauges are eventually used to evaluate the model performance, because they are flooded, active, and not influenced by water diversions (see **Table 4.1** for a detailed gauge summary).

### 4.4.2. Study domain and model configurations

#### *4.4.2.1. LSM Input Datasets*

The modeling domain is then set up for the SJRB, the most impacted region encompassing the Houston metropolitan area (**Fig. 4.6**). This basin has a drainage area of 10,274 $km^2$ and an annual rainfall of 1100–1400 mm, which drains into the Galveston Bay. Its vicinity to the Gulf of Mexico makes it vulnerable to damages and economic losses from hurricane and storm surges of the Atlantic origins (Keim et al. 2007). **Fig. 4.6b** shows the domain land cover types as generated from the WRF pre-processing tool (WPS, http://www2.mmm.ucar.edu/wrf/users/wpsv3.8/). WPS archives a database of land cover type, soil texture category, and vegetation greenness for the entire globe, and the domain-specific input datasets are automatically created after the grid resolution, geographic bounding box latitudes/longitudes information, and map projections are specified for the WPS. The land cover data for the default WPS v3.8 are from the 20-category MODIS data, which better captured the most recent urban expansions of the Houston metropolitan area compared to the 24-category USGS land cover data for the default WPS v3.7 (as validated against the Google Earth Professional and ESRI Basemap). Thus, the 20-category MODIS land cover data (**Fig. 4.6b**) are used in this study.

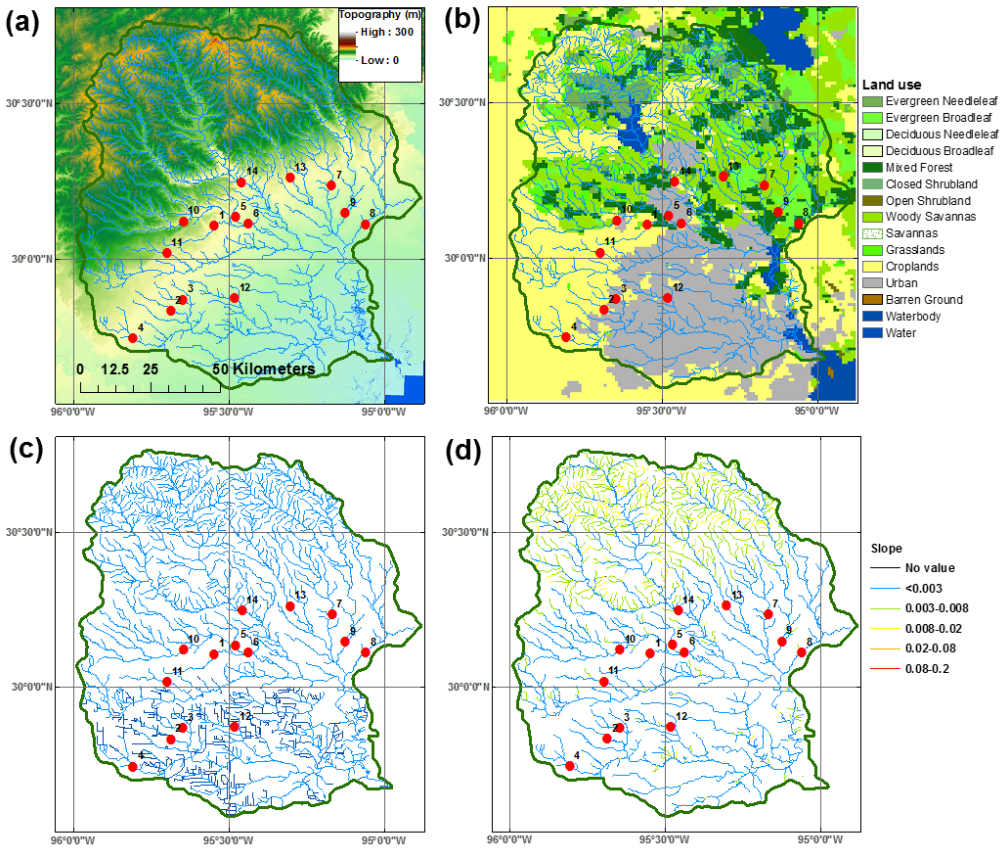


**Figure 4. 6.** The WRF-Hydro-RAPID modeling domain for the SJRB: (a) shows topography (meters) based on the 30-m National Elevation Dataset, (b) shows the land cover types in the 1-km modeling experiments, (c) shows all NHDPlus flowlines including artificial flow paths (dark blue) and natural drainage lines (light blue), and (d) shows the channel slope for the 1916 natural drainage lines.

### *4.4.2.2. River Network Input Datasets*

Several input files are required to route water horizontally with WRF-Hydro-RAPID. These include (1) a file documenting the river network's topological connectivity, (2) a file describing the coupling interface, and (3) two files defining the Muskingum routing parameters (k and x) for all NHDPlus river reaches. The SJRB contains 2332 NHDPlus river reaches (**Fig. 4.6c**), with 1916 being natural drainage lines (light blue) and 416 being artificial water pathways such as pipelines and ditches (dark blue). While these 416 artificial water pathways hold promise for future incorporation of urban drainage networks, such a model development is beyond the scope of this

current study, and thus we only consider 1916 natural lines with flow directions for the routing (**Fig. 4.6d**). Each flowline is associated with one local catchment areas, and several geospatial attributes including the channel slope (**Fig. 4.6d**) and mean annual flows.

An ArcGIS python toolbox (https://github.com/Esri/python-toolbox-for-rapid) is now publicly available and is used to generate the river connectivity file and the file for describing the area-weighted coupling interface. ArcGIS functions ("feature to point" and "extract values to point") are used to obtain a file describing the centroid-based coupling interface (see an example of this procedure when RAPID was loosely coupled to a coarse-resolution LSM: http://rapid-hub.org/docs/RAPID_LSM_coupling_procedure.pdf). As for the Muskingum routing parameters, an optimal set of k and x derived from David et al. (2013) is used. David et al. (2013) derived this optimal set of routing parameters by using hundreds of USGS gauge observations in an 8-yr experiment from 2000 to 2007 (see more discussions in Section 4.5). In our study domain, there is a major reservoir (Lake Conroe, 85 $km^2$) located at the main stream of the San Jacinto river ~20 km upstream of the evaluation site Gauge #14 (see **Fig. 4.6b**, USGS gauge ID 08068000). To account for the possible reservoir flood controls that can influence the simulation results, we force the RAPID model with daily observed streamflow at the exact location of the dam, and the forcing is only applied to the particular river reach where the dam is located.

#### *4.4.2.3. Model Configurations*

WRF-Hydro is configured in an "offline" mode to conduct the modeling experiments. The Noah-MP LSM is used as the vertical column model here while future studies could take advantage of the multiple LSM feature of WRF-Hydro to account for uncertainties with different hydrological models (rather than uncertainties with the parameters or parameterizations in a single model). Noah-MP supports multiple physical parameterization schemes, including the free-drainage (FD) and the simple groundwater-based (SIMGM) options for runoff parameterization. SIMGM is more physically-complete that relates groundwater with runoff using exponential-decay functions (Niu

et al. 2007), yet it is not used in this study for two reasons. First, SIMGM was previously evaluated in global major river basins (Niu et al. 2011) but not in smaller basins where complex surface water–groundwater interactions might exist. Second, we find the SIMGM-simulated flood hydrographs for this particular event are slightly worse than that by FD, which may be related to the inappropriate subsurface parameters used by SIMGM. Therefore, to avoid discussions on the SIMGM groundwater representations and to keep it the same as the operational NWM, the FD runoff parameterization (Schaake et al. 1996) is consistently used in this study. To simulate the discharge from September 12 to 17, the model is initialized on August 1 to allow for more than one month of spin-up. The soil moisture outputs from the Noah model of the North American Land Data Assimilation System (NLDAS-2) (Xia et al. 2012) are interpolated and used for the initialization on August 1. In general, Gochis et al. (2015) suggested a longer spin-up time for hydrological applications with WRF-Hydro, yet we consider the 1.5-month spin-up as adequate in this case for the following reasons. First, our results with the RAPID routing are minimally influenced in a one-year spin-up experiment for this flood. Second, the grid-based routing in WRF-Hydro has a dependence on the surface water head on the channel grids, while RAPID does not (**Fig. 4.2**); this may make the spin-up time with RAPID quicker than that with the grid-based routing. Third, for big floods with significant amount of surface runoff, the model sensitivity to spin-up time might be intrinsically less. However, for flood events with less surface runoff, longer spin-up time might still be needed. In line with Cunha et al. (2012) and Yilmaz et al. (2008), we do not calibrate the LSM parameters aiming to better understand the model behavior through "calibration-free" model evaluations. Thus, the default Noah-MP parameters adopted by previous large-scale studies (Cai et al. 2014; Niu et al. 2011; Yang et al. 2011) are used. For all experiments (**Table 4.2**), StageIV is used as the rainfall forcing, and the NLDAS-2 provides other forcing

variables such as the wind, temperature, humidity, and pressure. The temporal resolution is 1-hr for the LSM and 15-min for RAPID (David et al. 2013). Exps. 1–4 are conducted to assess the hybrid model's sensitivity to the LSM grid resolution (i.e. 4 km and 1 km) and the grid-to-vector coupling interface (i.e. a centroid-based versus an area-weighted approach).

| | **Land Resolution** | **Grid-to-Vector Coupling Interface** | **Muskingum Routing *k*** |
|---|---|---|---|
| Exp1 | 4 km | Centroid-based, Fig. 4.3b | Optimized by David et al. (2013) |
| Exp2 | 4 km | Area-weighted, Fig. 4.3c | Same as above |
| Exp3 | 1 km | Centroid-based, Fig. 4.3d | Same as above |
| Exp4 | 1 km | Area-weighted, Fig. 4.3e | Same as above |

**Table 4. 2**. Modeling experiments conducted in this study.

## 4.5. RESULTS

### 4.5.1. Resolution and coupling interface

**Fig. 4.7** shows the model-simulated flood discharges compared against observations at the 14 USGS gauges. Overall, the model can simulate the flood discharge reasonably well at all these gauges with contributing drainage areas ranging from 55.68 $km^2$ to 2144.51 $km^2$, despite what LSM grid resolutions and the coupling interfaces were used. Considering half of these gauges have a contributing area of less than 200 $km^2$ and locate in small creeks with mean annual flow less than 5 m3 s-1, these results seem quite promising because flood forecasting for smaller drainage basins remains a challenging task for many hydrologic models. Regionally, the model also tends to perform better at gauges with relatively faster rate of rise in more urban environments (i.e. steeper rising limbs, see locations with COMID 1508045, 1508279, 1440311, 1440237, 1440211, and 1438725), which has an average NSE of 0.63±0.15. In comparison, the model performs slightly

worse at those with slower rising limbs (see locations with COMID 1520249, 1520237, 1520083, 1508047, 1520091, 1508121, and 1508043), where the average NSE is lower at –0.54±1.05.

To more comprehensively assess the regional model performance, the following skill metrics were used to quantitatively assess the model performance, which include (1) Nash-Sutcliffe Efficiency (NSE $=1-\sum_i^n(M_i-O_i)/\sum_i^n(O_i-\bar{O})$), (2) Pearson correlation coefficient (CC $=cov(M,O)/(\sigma_M\cdot\sigma_O)$ ), (3) percentage bias in accumulated discharge (PBIAS $=(\sum_i^n M_i-\sum_i^n O_i)/\sum_i^n O_i\times 100\%$ ), and (4) normalized root-mean-square-error (NRMSE $=\sqrt{\sum_i^n(M_i-O_i)^2/n}/\bar{O}$), where $M_i$ and $O_i$ represent hourly river discharge from the model and the observation, respectively (n hours in total). These skill metrics are commonly used in evaluating hydrological and environmental models (Gupta et al. 2009) from different perspectives. For example, NSE assesses the overall hydrograph, CC assesses co-variance of two time-series, PBIAS and NRMSE both assess the error magnitude in the runoff volume. Below, boxplots of these statistics are shown because it summarizes the regional performance with the best, 75$^{th}$, median, 25$^{th}$, and worst results.

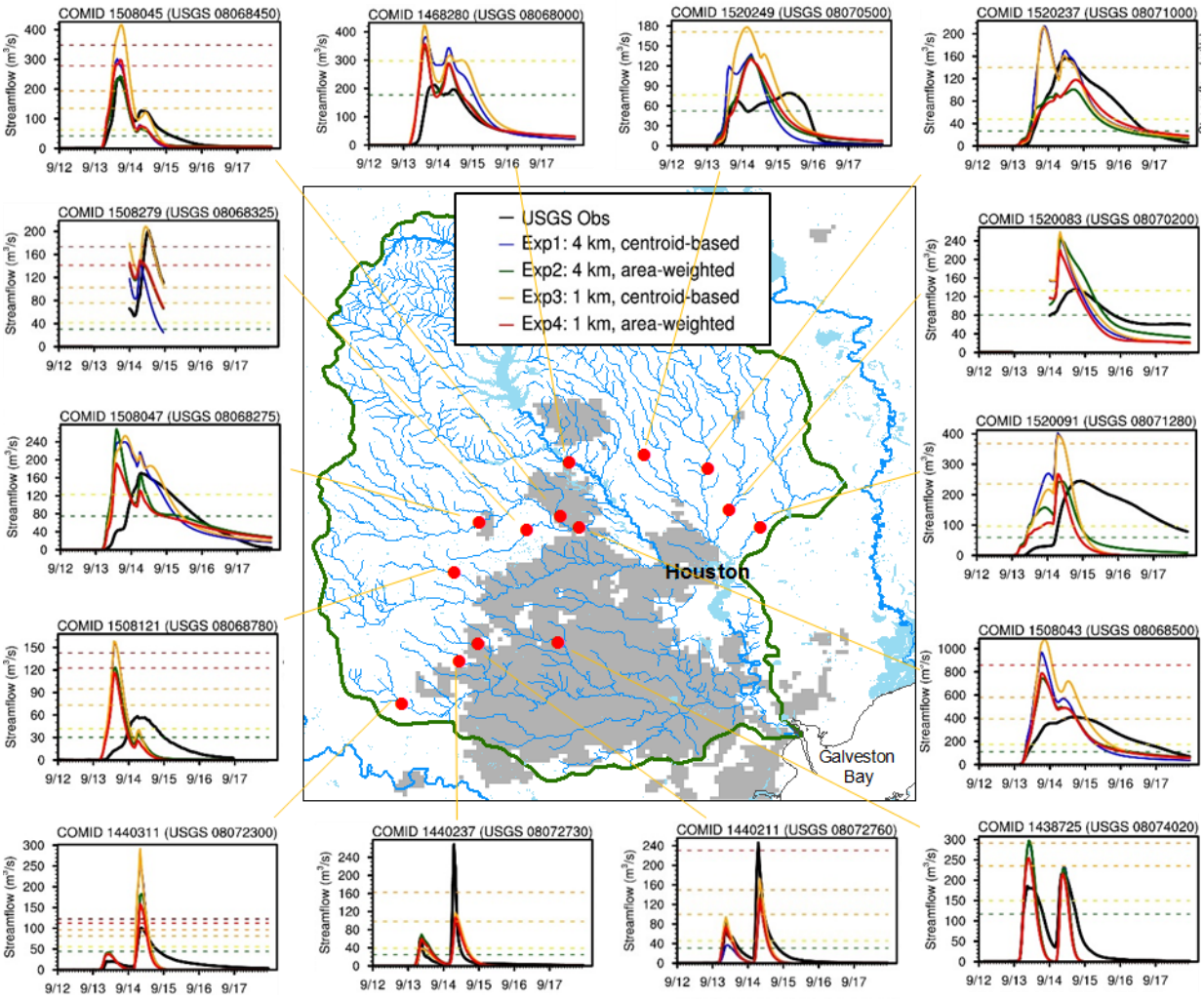


**Figure 4. 7.** Simulated flood hydrographs by Exps. 1–4 (color lines) and their comparisons with gauge observations (black lines) at 14 USGS gauge locations. Horizontal lines show the discharge at different flood return period levels using the same color legend in Fig. 4.5. Grey shaded areas are the urban land cover type.

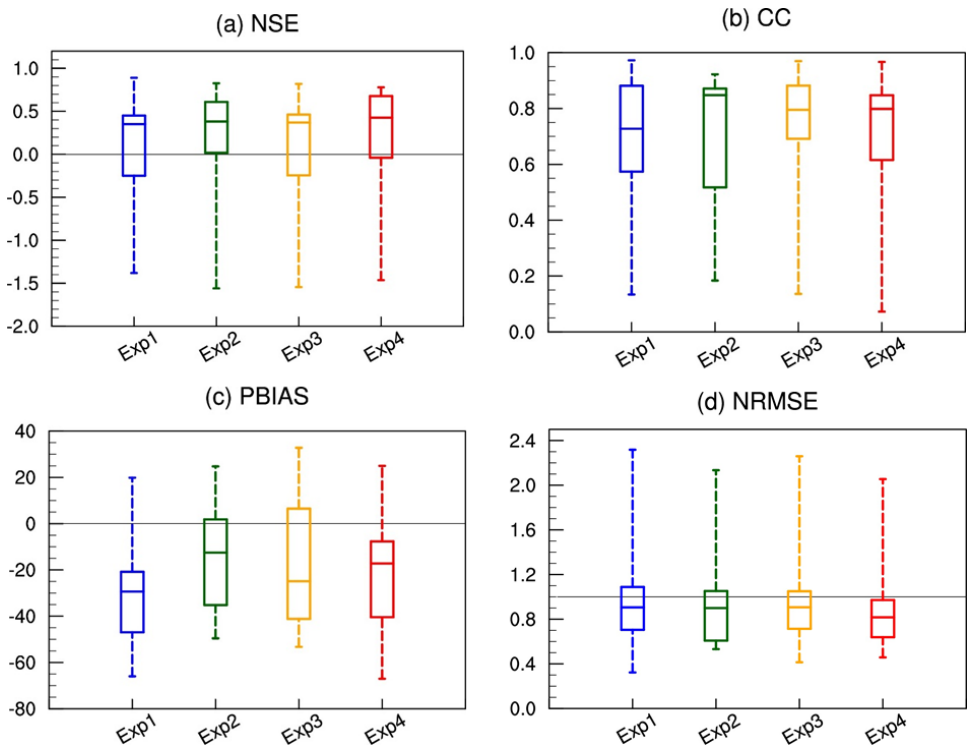


**Figure 4. 8.** Boxplots of the modeling experiment statistics comparing against 14 USGS gauge observations. Four statistics shown include (a) Nash-Sutcliffe Efficiency (NSE), (b) correlation coefficient (CC), (c) percentage bias (PBIAS) in accumulated streamflow (%), and (d) normalized root-mean-square-error. Statistics are calculated using hourly data from 0000 UTC 12 Sep to 0000 UTC 17 Sep 2008.

Among the four experiments, Exp4 consistently exhibits the highest NSE (median: 0.43) and the lowest NRMSE (median: 0.82 $m^3$ $s^{-1}$), meeting the expectation that the model performs better with a finer land grid resolution and an area-weighted coupling interface. Exhibiting slightly higher CC and lower PBIAS, Exp2 also shows slightly worse median NSE (0.38) and NRMSE (0.90 $m^3$ $s^{-1}$), which may be due to the coarser land grid resolution that influences the runoff generation through changing the land cover, greenness, and soil texture data accuracy. Note that difference between Exp2 and Exp4 may be also influenced by the lateral inflow calculations (Eqns. 1–2) resulting from the change in spatial correspondence between the grid and vector modeling units (**Fig. 4.3**), but the overall model sensitivity to these two configurations is small. Exp1 and Exp3 both exhibit lower median NSE and greater negative PBIAS due to a lack of accurate accounting of the lateral runoff contribution. However, it is noted that all the calibration-free (for

the LSM part) modeling experiments exhibit the highest NSE skill reaching 0.71 to 0.82, suggesting the flood responses at those certain gauges are not sensitive to the examined grid resolutions and the coupling interface. Using an area-weighted coupling interface (Exp2 and Exp4) tends to result in slightly narrower range in the $25^{th}$ to $75^{th}$ quartiles of these boxplots, suggesting more spatially-consistent model performance.

The hybrid model is more sensitive to the coupling interface than the grid resolution, as can be seen from the smaller difference between Exp1 and Exp3 (with different grid resolution) than that between Exp1 and Exp2 (with different coupling interface); the same conclusion holds true for Exp2 and Exp4 Exp3 exhibits greater range in PBIAS, which suggest it has more variable model performance across the spatial domain than other experiments, because Exp3 is implemented at high resolutions (1-km), thus it can produce very non-representative lateral inflows when the catchment size largely exceeds the grid cell size (see the example catchment in Fig. 4.3d and the 1km×1km centroid cell to pass runoff). For places whose catchment size is smaller than 1km×1km (see some small catchments in the bottom of Fig. 4.3d), however, the centroid cell's runoff may represent the lateral inflows well. The non-linear spatial correspondence between the irregular NHDPlus catchments and the regular LSM grid cells helps explain why Exp3 exhibits more variable PBIAS. In Exp1, because the grid cell size (16 $km^2$) exceeds the mean catchment area of the SJRB (5.23 $km^2$), runoff extracted from the 4km×4km centroid cell may be better able to represent the lateral inflow contributions in several occasions. For a hybrid modeling framework containing both regular (i.e. LSM grid cells) and irregular non-uniform modeling units (i.e. NHDPlus catchments), these results suggest if the LSM grid resolution is high, an area-weighted coupling is key in calculating representative lateral inflows; this coupling, however, may become less important with coarser-resolution LSM grids in which case a simpler centroid-based coupling may be justified, similar to previous studies performing the routing using the 1/8º LSM runoff (David et al. 2013).

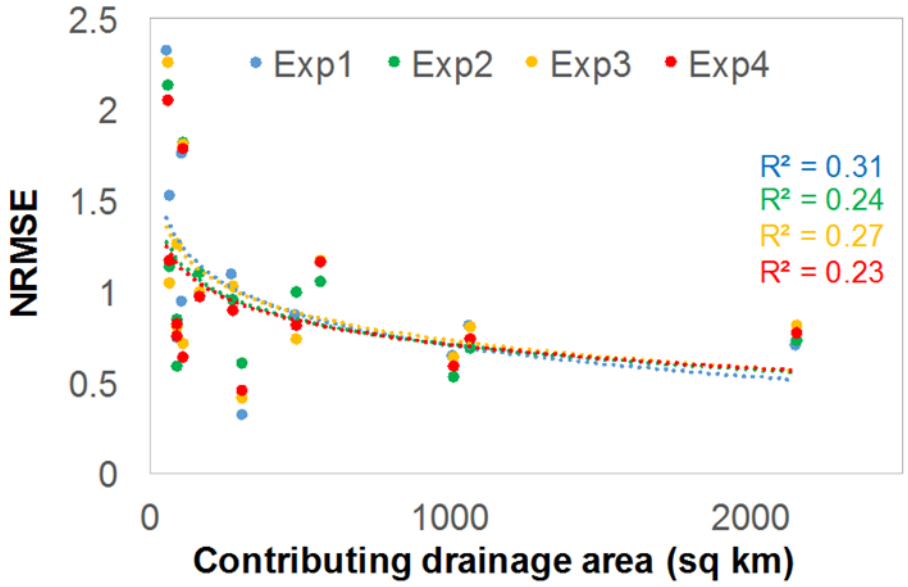


**Figure 4. 9** Scatter plot between the NRMSE and the contributing drainage area ($km^2$) of the 14 gauges. $R^2$ shows the coefficient of determination of the best fitted function.

**Fig. 4.9** further shows some interesting patterns when plotting the statistics of the 14 gauges against their contributing drainage areas. NRMSE is examined here because we expect the runoff volume simulation to be more influenced by drainage areas while the covariability of time series (i.e. CC) and the overall hydrograph (i.e. NSE) are potentially influenced more by other factors such as the travel time representation in the model. Also, NRMSE normalizes the runoff simulation error using the mean observed discharge. We see from **Fig. 4.9** that for all experiments, the NRMSE–drainage area relationship follows a power law function with coefficients of determination (R2) ranging from 0.23–0.31, which means that the model's ability to capture the right amount of runoff generally increases with contributing drainage areas. As suggested by Cunha et al. (2012), such a relationship is described as the "aggregation effect" of the river network, meaning that as a river goes downstream and aggregates a larger contributing area, the model errors resulted from various upstream locations tend to be cancelled out. Therefore, a river network routing model also tends to have less simulation uncertainties at downstream locations, the findings of which are in accordance with our results here. Interestingly, we also find that R2 for Exp1 and Exp3 is greater than that in Exp2 and Exp4. This suggests that the "aggregation effect"

is more obvious in experiments using a centroid-based coupling interface (Exp1 and Exp3) than that using an area-weighted coupling interface (Exp2 and Exp4). This is reasonable because in the latter coupling interface, the runoff volume can be better captured at both upstream and downstream gauges, thus, its dependence on the drainage basins to filter out errors becomes less. This provides another support to suggest the area-weighted coupling as a useful and optimal approach to reduce the "aggregation effect" in achieving better regional streamflow simulations.

### 4.5.2. GIS-based flow travel time estimation using NHDPlus

In addition to the grid resolution and vector-raster coupling, we next examine the flow travel time representation in the model to help explain the spatially-varied regional model performance. In the Muskingum method, the routing parameter k is interpreted as flow travel time that dominates the overall simulated hydrograph while x being much less influential (Kim et al. 2001). Although an optimal set of k and x obtained from David et al. (2013) was used, it is noted that these parameters were derived from a global optimization algorithm using daily discharge observations at hundreds of USGS gauges in an 8-year modeling experiment in Texas. Questions are then raised as regard to the validity of using these parameters to resolve the sub-daily flow concentration processes. In addition, the factor to minimize the least square error used in the optimization scheme was applied globally (see Eqn. 6 in David et al. (2013)), which increases or decreases in the same direction at different gauge locations. As a result, it is noted that their representativeness across a spatial domain also needs to be investigated.

To examine this issue, we adapt a GIS-based method to estimate spatially-distributed flow travel time using the NHDPlus geospatial data. Different from previous RAPID applications that adjust total travel time (k, in seconds) based either on optimization or spatial variability that separates larger river basins (e.g. David et al. 2011; Tavakoly et al. 2017), the flow travel time is estimated for each NHDPlus channel and catchment in this study to account for more detailed spatial variability. It is under the conceptualization that the LSM-generated runoff first travels

along the catchment plane with time $T_{ov}$, and is then routed along the channel network with time $T_{ch}$. From the GIS data structure perspective, $T_{ov}$ and $T_{ch}$ take place on the polygon features (catchments) and the line features (flowlines), respectively. This estimation can then be related to the NHDPlus geospatial attributes and land cover types, using equations (Eqns. 4.3 and 4.4) widely adopted in other hydrological models from the GIS modeling community (Cho and Engel 2017; Neitsch et al. 2011) that build upon the Hydrologic Response Unit (HRU). Similar to RAPID, these models use irregular modeling units that may be inherently better for hydrological applications than traditional grid-based LSM discretization. The difference is that NHDPlus is a pre-established data fabric with "local" catchments (i.e. the drainage area of the current river reach minus that of the upstream reach), thus it incorporates fine-scale features (from 30-m DEM, Section 4.4.3.2) whose average catchment size is 2–3 km$^2$.

$$T_{ch} = \frac{0.000172 \cdot L_{ch} \cdot {n_{ch}}^{0.75}}{area^{0.125} \cdot {slp_{ch}}^{0.375}} \quad \text{Eqn. (4.3)}$$

$$T_{ov} = \frac{419.3 \cdot {L_{ov}}^{0.6} \cdot {n_{cat}}^{0.6}}{{slp_{cat}}^{0.3} \cdot {P_{ov}}^{0.4}} \quad \text{Eqn. (4.4)}$$

In the above equations, $L_{ch}$, $n_{ch}$, and $slp_{ch}$ denote the flow length (m), Manning's roughness coefficient, and slope for the channel, respectively; while $L_{ov}$, $n_{cat}$, and $slp_{cat}$ denote the flow length (m), Manning's roughness coefficient, and slope for the overland catchment, respectively. $area$ is the sub-basin area in previous GIS-based models such as SWAT, and is replaced with the NHDPlus catchment area (km$^2$) in our case. The constants are used to convert the travel time to seconds. The $T_{ov}$ formulation assumes an average net incoming flux of 6.35 mm hr$^{-1}$, and we replace that with $P_{ov}$ (mm hr$^{-1}$) as the domain-average net incoming flux. Readers are directed to Neitsch et al. (2011) and Cho and Engel (2017) for more details on the equations.

To calculate $T_{ch}$ and $T_{ov}$, $L_{ch}$ and $slp_{ch}$ are readily available from the NHDPlus geospatial data attributes. To estimate the overland flow length, half length of the NHDPlus catchment area divided by the NHDPlus flowline length ($0.5 \cdot area \cdot 1000000/L_{ch}$) is used, which is similar to

Li et al. (2013). Channel roughness $n_{ch}$ and overland slope $slp_{cat}$ are set to 0.035 and 0.01 for the entire domain, respectively, because they are reasonable initial estimates for large-scale studies in the absence of further distributed data. Overland roughness coefficient $n_{cat}$ is a function of the land cover type, which has a one-on-one relationship based on the widely-used empirical values (see the lookup table "HYDRO.TBL" by Gochis et al. (2015)). For the net incoming flux $P_{ov}$, 7 mm $hr^{-1}$ is used based on the event rainfall analysis averaged over the SJRB (see **Fig. 4.4c**).

Based on the pre-processing in GIS, Tch and Tov (hrs) for the SJRB can be estimated as shown in **Fig. 4.10**. Across the spatial domain, river reaches with longer lengths and flatter channel slopes (**Fig. 4.6c**) generally have greater Tch (**Fig. 4.10a**). Similarly, catchments with rougher surfaces, for example forest in the northeast SJRB and cropland in the southwest SJRB (**Fig. 4.6b**), also tend to have longer Tov than other areas of the domain (**Fig. 4.10b**). Under this conceptualization, the total travel time for each NHDPlus modeling unit ($T_{total}$) can be represented by the sum of Tch and Tov. Comparisons between the probability distribution function (PDF) of the calculated $T_{total}$ and that optimized by David et al. (2013) suggest that the optimized travel time is shifted to the lower end of the PDF, while the calculated $T_{total}$ has more modeling units with longer travel time (**Fig. 4.10c**). This explains why the model performance in Exps. 1–4 tends to show peakier than observed hydrographs for gauges with forest and crop land cover types, while exhibiting good performance for those with faster rises (**Fig. 4.7**); such a problem was introduced by the optimization scheme that decreases the travel time k in the same direction for all gauges, where the calculated $T_{total}$ showed better spatial heterogeneity across the domain in this case.

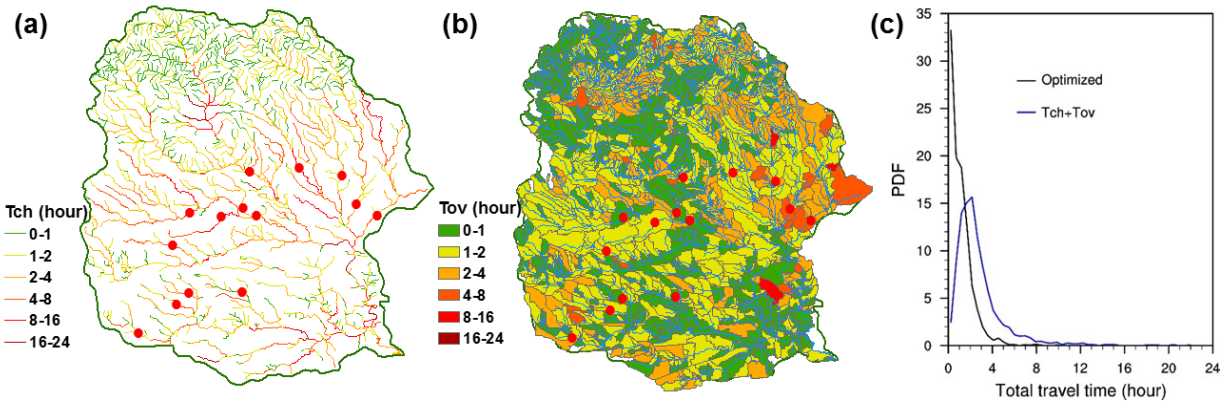


**Figure 4. 10.** GIS-based estimation of horizontal water travel time using the NHDPlus geospatial datasets: (a) channel travel time, Tch and (b) overland travel time, Tov. (c) shows the comparison between the probability distribution function (PDF) of the total travel time (black) optimized using daily USGS observations by David et al. (2013) and that of the total travel time (blue) implemented in this study (Ttotal=Tch+Tov).

To test the sensitivity of the regional discharge simulation to travel time estimation, we conduct two additional experiments by replacing k with the $T_{total}$ calculated above. **Fig. 4.11** shows the NSE boxplots for the modeling experiments with optimized travel time k and the calculated $T_{total}$. In both the 4-km and 1-km modeling experiments, the highest NSE is achieved when an area-weighted coupling interface and the calculated $T_{total}$ are adopted (see **Fig. 4.11**). The regional median NSE is brought up from ~0.4 in Exp2 and Exp4 to ~0.65 for Exp5 and Exp6, which is due to the calculated $T_{total}$ that slows down the flow concentration processes at a range of locations experiencing moderate rate of rise floods (**Fig. 4.7**). However, we note that this adopted calculation only considered spatial heterogeneity in estimating flow travel time. Future research is warranted to further introduce temporal variability to account for the dynamical flow travel time responses, because it would be better adapted to actual flood prediction scenarios in which the routing parameters cannot be obtained prior to these particular events.

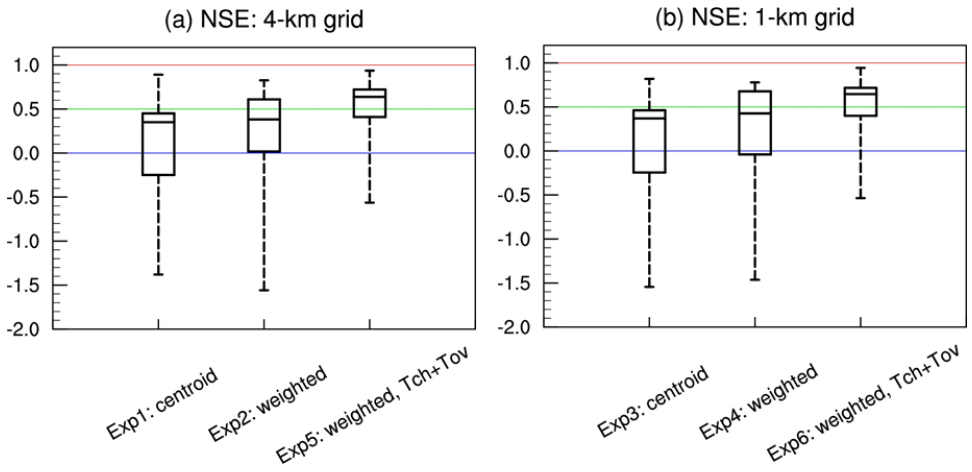


**Figure 4. 11.** The NSE for modeling experiments with optimized travel time k and calculated total travel time (Tch+Tov). (a) and (b) show results from the 4-km and 1-km experiments, respectively. Blue, green, and red reference lines show NSEs of 0, 0.5, and 1, respectively.

## 4.6. DISCUSSIONS

### 4.6.1. Computational efficiency, model complexity, and applicable settings

**Fig. 4.12** shows an independent test of the model computational costs based on several experimental runs for the Texas Region 12. This domain has a drainage area of 464,135 $km^2$ and 68,143 NHDPlus river reaches, and the model is configured at 4.5-km grid resolution for a 3-day simulation in this test. Using the TACC Stampede supercomputing facility, it is shown that RAPID took ~5% of the LSM simulation time with streamflow outputs recorded at all 68,143 river reaches, which highlights the computational efficiency of this hybrid model when it is applied to a large spatial domain. The best computational performance is achieved using 8 to 16 computing cores.

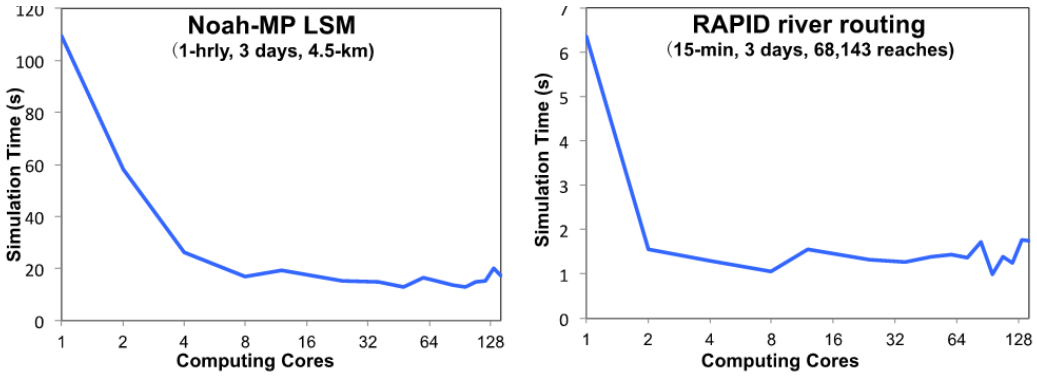


**Figure 4. 12.** Computational time using multiple cores for a 3-day simulation for the Texas Hydrologic Region 12 with 68,143 NHDPlus river reaches. Left and right panels show the cost for the Noah-MP and RAPID components, respectively.

| | **Grid-based** | **Vector-based (RAPID)** |
|---|---|---|
| **Routing time step (suggested)** | 15s<br>(Gochis et al. 2015) | 900s<br>(David et al. 2011b) |
| **Routing computational time (single core)** | 3212.66s<br>(Outputs configured for 14 user-specified gauge locations) | 14.67s<br>(Outputs available at all NHDPlus river reaches) |
| **Explicitly resolving overland flow** | Yes | No |
| **Feedback to the atmosphere** | Yes | No |
| **Applicable settings** | Coupled and offline hydrometeorological applications | Offline hydrological and GIS applications |

**Table 4. 3**. Comparisons between the 250-m grid-based routing and the vector-based routing based on a modeling experiment for the SJRB using the 4-km LSM grid.

We also compare the full-feature grid-based routing (hereafter FFGR, based on the 250-m grid) with the RAPID routing in **Table 4.3** (both are currently supported in WRF-Hydro) to facilitate a discussion on the model efficiency, complexity, and their applicable settings. In FFGR, we turn on the saturated subsurface routing, overland surface routing, channel routing (option 3: diffusive routing), and groundwater baseflow bucket (option 3: pass through) to consider all processes represented in the grid-based routing. RAPID, on the other hand, simplifies the explicit

representation of these processes and only focuses on the streamflow response at the reach scale as an end product of the travel time estimation. In the SJRB, FFGR and RAPID respectively takes 3212.66 s and 14.67 s for routing, the efficiency of the latter is partly due to the object-oriented feature of RAPID and partly due to its coarser time step tied to the average length of the modeling unit, i.e., 250 m for FFGR and 2470 m for RAPID (see more conceptual discussions in Section 4.3). The model performance of the uncalibrated FFGR (median NSE: 0.31) is slightly worse than RAPID using optimized routing parameters with centroid-based coupling (median NSE: 0.35; Exp1) and that with area-weighted coupling (median NSE: 0.38; Exp2). We note that the NSE comparison here is not to discuss the two routing approaches in terms of their skill in simulating or prediction floods, because pronounced differences still exist in their degrees of calibration and the diffusive/kinematic wave approximations in addition to differences in the modeling units. The results only show that RAPID can serve as a computationally efficient alternative for "offline" hydrological applications if carefully chosen (or physically-based) routing parameters are given. The FFGR, on the other hand, is more physically-complete in explicitly representing the flow paths and moisture redistribution using a grid-by-grid routing approach, which may be suggested for studies to understand detailed water movements and/or their feedbacks to the atmosphere at a higher computational cost.

### 4.6.2. Model efficiency and complexity for operational use

The considerations on efficiency and complexity are also seen in different configurations of the NWM, the new operational hydrologic prediction tool for the United States using WRF-Hydro as the backbone. Building upon the Noah-MP LSM and the channel routing on the 2.7 million NHDPlus river reaches (Maidment 2017), the NWM is now producing real-time predictions for streamflow and all LSM-simulated states and fluxes at 18-hr (short-range), 10-day (medium-range), and 30-day (long-range) forecasts, together with one historical/analysis run to support the "water-intelligence" of the nation (http://water.noaa.gov/about/nwm). While the short-

to medium-range NWM configurations implement the FFGR routing to account for the best-supported physical representations (albeit needing further performance evaluation and parameter calibration), the long-range forecast implements a scheme similar to the presented WRF-Hydro-RAPID framework because of trade-off considerations between affordable computational power and model complexities. Thus, the hybrid modeling strategy and results for flow travel time estimation should inform the NWM long-range forecast, which is expected to improve upon those studies only separating the flow travel time estimations for larger river basins or using different river reach lengths. However, future research is warranted to introduce temporal variability (Du et al. 2009) to better adapt these parameters for use in actual predictions.

Several geospatial datasets with NHDPlus are also discussed here for potential future implementation in the hybrid framework, including artificial water pathways such as pipelines and ditches (Section 4.3), small- to medium-sized lake ponds and swamps, dams, springs, and other scattered features on land (not shown). These urban drainages (line feature), lake ponds (polygon feature), and dams and springs (point feature) are typically well-established and compiled by GIS as georeferenced data fabrics (Lehner and Grill 2013; Lin et al. 2015) at the continental to global scales. However, these features are difficult to be systematically represented in traditional grid-based LSMs due to their geographically-scattered characteristics. The presented hybrid framework has the potential to explicitly incorporate them from the perspective of flexible data structures (Mizukami et al. 2016), but future work engaging the GIS modeling with the LSM developments is needed to take better advantages of these vector-based data.

## 4.7. CONCLUSIONS

A vector-based river routing model named RAPID was implemented as a new routing option in the community WRF-Hydro modeling system to facilitate large-scale hydrological modeling. This hybrid model combines the grid-based and vector-based modeling units, which was demonstrated to be computation ally efficient, currently suitable for use in “offline”

hydrological applications (see Appendix for the software availability). Such a model coupling work is in accordance with the need to advance land surface modeling and river routing by taking advantage of GIS-based modeling features (Clark et al. 2015; Li et al. 2013).

By conducting a case study to simulate the inland riverine flood during Hurricane Ike in 2008, we assess the land grid resolution and the grid-to-vector coupling interface in influencing the model performance. We show that the model has reasonable performance in predicting the inland flood discharge at 14 USGS gauges in a highly urbanized domain. The hybrid model is more sensitive to the grid-to-vector coupling interface than to the LSM grid resolutions, and a 1-km land grid and an area-weighted coupling interface is suggested to show the optimal model performance. Note that we consistently use the 4-km rainfall input in this study, and it is possible that the resolution inherent to the rainfall input can overwhelm the model's sensitivity to the examined grid resolutions. However here we only focus on the grid resolution changes without imposing changes to the rainfall input's native resolution. Future studies using the new operational radar rainfall, i.e., the Multi-Sensor Multi-Radar (MRMS) at 1-km resolution, may be warranted to further examine this issue. A GIS-based approach to estimate flow travel time can be adapted to further improve the regional model simulation skill. Results presented in this study have implications for the NWM, whose operational configurations consider the trade-offs between the model efficiency and complexity. The hybrid modeling framework holds promise for modeling other irregular hydrologic features, yet future work remains to be done in taking advantage of the under-exploited GIS geospatial data and introducing temporal variability for the flow travel time estimation, to further advance the current modeling capability presented in this study.

## 4.8. APPENDIX: SOFTWARE AVAILABILITY

The preliminary code package for the hybrid WRF-Hydro-RAPID model is downloadable at https://www.ral.ucar.edu/sites/default/files/public/projects/wrf_hydro/v3_0/. More updated versions are available upon request to the corresponding author (liang@jsg.utexas.edu). The

updated code package uses RAPID v1.3.0 (David et al. 2013b) that uses hash table features for efficient river network initializations. Model compilation needs installation of NetCDF and PETSc for model I/O and matrix-based computation (suggested versions: netcdf/3.6.3 and petsc/3.4). Environment variables (WRF_HYDRO, WRF_HYDRO_RAPID) need to be set to 1, and successful compilation will generate an executable 'wrf_hydro.exe'. In the 'hydro_namelist', specify the channel routing option as 4 for running RAPID. More details can be found at https://github.com/c-h-david/rapid and http://rapid-hub.org/docs/RAPID-WRF-Hydro_install.pdf.

## 4.9. ACKNOWLEDGEMENT

This work is supported in part by the National Natural Science Foundation of China under grant number 41375088, and in part by the Microsoft Research and the Jackson School of Geoscience, UT-Austin. Cédric H. David is supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. David Gochis and Wei Yu are supported by the National Science Foundation through its cooperative funding of the National Center for Atmospheric Research (NCAR). Additional support for Gochis and Yu were provided by NSF EarthCube Grant #1343811. Kevin Sampson (NCAR) is acknowledged in providing GIS support. P.L., Z.-L.Y., D.J.G., D.R.M., and C.H.D. proposed the implementation of a vector-based river network model in the WRF-Hydro framework; P.L. worked on the code development with contributions from W.Y., M.A.S.-V., and C.H.D.; P.L. conducted the modeling experiments with inputs from Z.-L. Y. and D.J. G.

# Chapter 5: Insights into hydrometeorological factors constraining flood prediction skill during the May and October 2015 Texas Hill Country floods[2]

## 5.1. ABSTRACT

This study evaluates the May and October 2015 flood prediction skill of a physically-based model resembling the United States National Water Model (NWM) over the Texas Hill Country. It also probes into factors contributing to a record flood at Blanco River at Wimberley (WMBT2) in May 2015. Using three radar-based quantitative precipitation estimation (QPE) products from Stage IV and Multi-Sensor Multi-Radar (MRMS) as inputs, it is shown that the event precipitation accuracy dominates the prediction skill, and that the finer-resolution MRMS QPE mainly benefits basins with small drainage areas. Overall, the model exhibits good performance at gauges with fast flood response from causative rainfall and gauges that are not forecast points in the National Weather Service's (NWS) Advanced Hydrometeorological Prediction System, showing great promise for forecasts, warnings, and emergency response. However, the model suffers from poor prediction skill over regions without rapid flood response and regions with human-altered flows, suggesting the need to revisit the channel routing algorithm and incorporate modules to represent human alterations. Two contrasting flood events at WMBT2 with similar meteorological characteristics are examined in greater detail, revealing that the location of intense rainfall combined with land physiographic features are the key factors for the flood response differences. Model sensitivity tests further show the record flood peak could be better obtained by tuning the initial wetness in deeper soil layers and the flow velocity field in the river network. These results offer hydrometeorological insights into the causes for such a record flood and why the model struggles to predict the record flood peak. (KEYWORD: Flash Flood Alley; Radar QPE; National Water Model; May and October 2015 central Texas floods; record Wimberley flood)

[2] This chapter was previously submitted to *Journal of Hydrometeorology*.

## 5.2. INTRODUCTION AND MOTIVATION

Improving forecasts and impact-based decision support for flooding is critical because it is the deadliest severe weather hazard in the United States, with a 10-yr average of over 90 fatalities a year (NWS 2016a). Flood fatality and damage costs continue to increase (Downton et al. 2005; Pielke et al. 2002) as extreme rainfall events become more intense and frequent in the extratropics (Boucher et al. 2013; Trenberth et al. 2003). In flood and hurricane-prone Texas, climate model runs and analyses show the annual probability of a hurricane producing over 500 mm of rain increased from 1% during 1981–2000 to 6% in 2017, with a predicted increase to 18% by the end of the 21st century (Emanuel 2017). Recent catastrophic heavy rainfall and flood events associated with historic pluvial periods like Spring of 2015 in the southern Plains and multi-day cases like Hurricane Harvey in August 2017 further underscore the need for better understanding and improving hydrological modeling capabilities.

Improving hydrological prediction is of vital importance to the Texas Hill Country, also known as "Flash Flood Alley" because it is one of the most flash flood-prone regions of the contiguous United States (CONUS) (Saharia et al. 2017). Intense rain rates frequently occur along the Balcones Escarpment, a limestone terrain uplift separating the ~500 m higher Edwards Plateau to the west from coastal plains flowing into the Gulf of Mexico to the southeast (**Fig. 5.1b**). The region's vulnerability to simultaneous Gulf and Pacific moisture sources, efficient warm rain processes, and quasi-stationary mesoscale forcing mechanisms enhance orographic lifting to produce sporadic 6–24 h rainfall totals of 250–500 mm. Steep river channel beds and thin soils in relatively small basins with high stream velocities and runoff potentials contribute to rapid river rises (Furl et al. 2018; Furl et al. 2015; Sharif et al. 2010). In addition to the region's complex terrain and soil characteristics, increased urbanization in the Hill Country and Austin–San Antonio corridor results in even faster river rises and shorter lag times between heavy rainfall and peak discharge than in the downstream coastal plains and most parts of the United States (Leopold 1991; Smith et al. 2000). These aforementioned factors combined with hundreds of low water crossings

and rapid population growth, contribute to significant fatality rates caused by regular short return-period events instead of rare, high-casualty events in other parts of the United States (Ashley and Ashley 2008).

Existing operational hydrological models struggle with accurately simulating the Hill Country's rapid flood response to heavy rainfall, limiting the region's flood preparedness. The lumped-parameter Sacramento Soil Moisture Accounting Model (SAC-SMA) does not account for spatial and temporal rainfall variations, and a distributed version of the model (Smith et al. 2012) also struggles to produce effective forecasts for small catchments and areas with complex terrain (Zhang et al. 2012). The launch of the National Atmospheric and Oceanographic Administration (NOAA) National Water Model (NWM; http://water.noaa.gov/about/nwm) represents an advanced conceptual shift in distributed hydrologic modeling using an earth system modeling approach (Sivapalan 2018). The NWM leverages the Weather Research and Forecasting Model Hydrological modeling system's (WRF-Hydro) (Gochis et al. 2015) modular structure to couple the Noah land surface model (LSM) with multi-parameterization (Noah-MP) (Niu et al. 2011) with atmospheric models and hydrological routing, which produces real time flood prediction at over 2.67 million locations (Maidment 2017). Although this represents a significant increase from around 4,000 forecast points currently with the NWS Advanced Hydrologic Prediction Service (AHPS; http://water.weather.gov/ahps/about/about.php) using the SAC-SMA model, the skill of the NWM and similar model configurations during extreme rainfall events is largely unknown.

The primary objective of this study is to evaluate how well a physically based model whose configuration resembles the NWM performs at the regional and local scale in the Hill Country during high-impact flood events. Using two seamless radar-based quantitative precipitation estimation (QPE) products, this study also assesses how QPE accuracy influences streamflow prediction and identifies the model's strengths and areas for potential improvements with the best observational constraints. Two heavy rainfall events on 23–24 May 2015 (hereafter the "May event") and 30 October 2015 (hereafter the "October event") whose annual exceedance

probabilities[3] were both less than 0.2% (1/500; NWS 2015, 2016b) are investigated in this study. The two events produced widespread flooding with record floods at the Blanco River at Wimberley (May event) and Onion Creek at Driftwood (October event), both of which were preceded by very saturated antecedent soil moisture conditions that helped intensify the magnitude and rapid response of the flooding. Multiple "Flash Flood Emergencies" were issued by the NWS Austin/San Antonio Weather Forecast Office (WFO) during both events due to numerous evacuations and high water rescues, along with 13 fatalities in the May event and five fatalities in the October event.

This study also attempts to offer mechanistic insights into the hydrometeorological drivers and basin physiographic features contributing to the Blanco River's record flood on 24 May 2015, whose peak flood stage of 13.7 m exceeded the last record on 28 May 1929 by 3.5 m. These insights are facilitated by comparing the model's relatively poor local performance in the May event to that occurred in October. Identifying what factors are missing in accurately predicting a flood event substantially exceeding its previous record should help drive improvements for the NWM and similar model configurations as presented in this study. These insights should also provide a better understanding of the complex nature of riverine flash floods in the Hill Country and regions with similar challenges to improve flood preparedness and resilience to future record flood events.

Note that this chapter focuses on assessing the hydrologic modeling part of the NWM framework by using the best available operational radar rainfall products as the forcing inputs. Therefore, the results presented here should be viewed as "simulations" instead of "predictions" that use weather forecasts as the model forcing inputs. Subsequently, any attempts to extrapolate the insights gained in this study to actual prediction scenarios need to be made with cautions. It is highly unlikely for the accuracy of weather forecasts, as those used in the operational NWM running at real time, to match that of the radar rainfall as used in this study. Hence, it is also

[3] Annual exceedance probabilities for Texas are currently based on Hershfield (1961), but preliminary updates to precipitation frequency estimates for Texas from *NOAA Atlas 14, Volume 11 Version 1* that will be finalized and published later in 2018 are available at https://hdsc.nws.noaa.gov/hdsc/pfds/PFDS-Texas/tx_peer_rev.html.

possible that the uncertainty of weather forecasts may overwhelm those originated from the hydrologic modeling. Nevertheless, it is the primary goal of this chapter to put more emphasis on the hydrologic modeling aspect of the NWM to reveal more potentially overlooked complexities associated with flood forecasting, in addition to weather forecasts. These efforts are expected to improve our understanding on the complex nature of floods, especially those ones exceeding historical record levels, which will need joint efforts not only from improved weather forecasts but also from better hydrologic and hydraulic representations of a physically-based model.

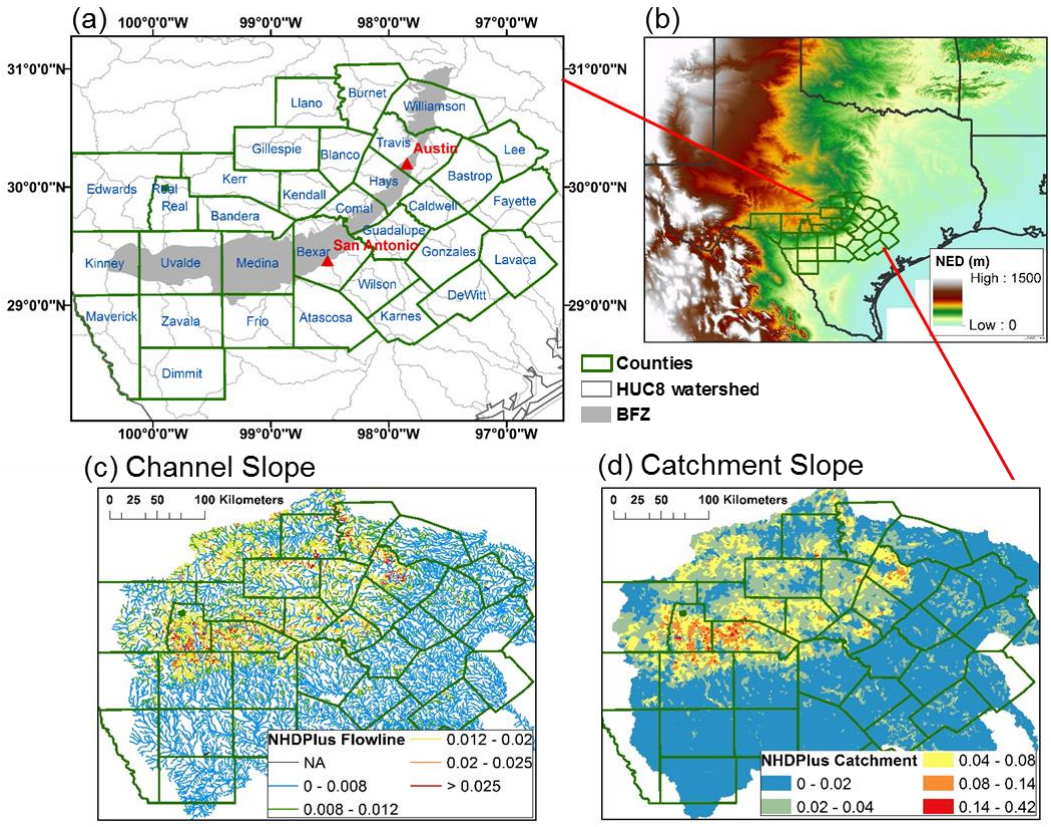


**Figure 5. 1.** The modeling domain in south central Texas. (a) shows the 33 counties and the Balcones Fault Zone (BFZ) for the location of the Balcones Escarpment. (b) shows the terrain (m). (c–d) show the channel and catchment slope for the 16,906 NHDPlus modeling units, respectively.

## 5.3. METEOROLOGICAL DESCRIPTIONS FOR THE MAY AND OCTOBER EVENTS

Both flood events exhibit broad meteorological characteristics that have been observed during past heavy rainfall events in the Hill Country. Grice and Maddox (1983) found heavy rainfall occurs most often in late spring or fall with quasi-stationary outflow boundaries left behind by previous convection similar to the May event [i.e., "mesohigh" patterns in Maddox et al. (1979)] or slow-moving "frontal" patterns like the October event. In both events, longwave 500–700 hPa troughs over the western United States digging into Mexico with a leading shortwave moving through Texas enhanced convection along mesoscale boundaries. A 7-yr (2009–15) local study showed these "longwave mesoscale" patterns produce the greatest areal coverage of heavy rainfall, resulting in the highest storm totals of any percentile compared to singular shortwaves or cutoff lows with or without synoptic fronts (Hopper and Hampshire 2016). Only four of these "longwave mesoscale" cases were identified out of 58 events analyzed by Hopper and Hampshire (2016), two of which are investigated in this study.

Similar mesoscale environmental parameters in both events included several ingredients for heavy rainfall. Precipitable water values (48–53 mm) were over two-standard deviations above the mean with a south–southeasterly 20 m $s^{-1}$ low-level jet transporting this moisture nearly orthogonal to the Balcones Escarpment. Relatively low lifting condensation levels and high –10ºC levels allowed for 5.0–5.5 km warm cloud layer depths (Schroeder et al. 2016), which enhanced efficient warm rain processes. Despite moderate shear, westerly meso-beta scale convective elements (Corfidi et al. 1996) that were slower and nearly perpendicular to the mean winds supported "echo training", back-building, and quasi-stationary convective modes (Schumacher and Johnson 2005, 2009), producing heavy rainfall in both events.

Quasi-stationary frontal forcing focused heavy rain in the October event, whereas outflow boundaries and a static mesoscale convective vortex (MCV) concentrated rainfall during the May event that Furl et al. (2018) described in greater detail. During the May event, convection that developed along an outflow boundary in the Hill Country and a sea breeze along the Gulf Coast

(**Fig. 5.2a**) merged over Kendall and southern Blanco Counties between 2000–2100 UTC 23 May 2015 (**Fig. 5.2b**). A quasi-stationary MCV that developed here by 2100 UTC encouraged cell mergers to maintain 51–102 mm h-1 (2–4 in $h^{-1}$) rain rates near the MCV through 0100 UTC 24 May 2015 before combining with a line of storms that moved in from Mexico (**Figs. 5.2c–5.2d**). One final convective line that developed along the west Texas dryline progressed across the region from 0300–1000 UTC (**Figs. 5.2e–5.2f**). During the October event, convection developed between 0600–0800 UTC 30 October 2015 (**Fig. 5.3a**) over the Rio Grande Plains, intensifying as it moved east into a strengthening low-level jet in the Hill Country through 1000 UTC (**Fig. 5.3b**). Discrete supercells that developed along the warm front fed a back-building, quasi-stationary line of convection from 1100–1530 UTC (**Fig. 5.3c–5.3d**). These convective echoes trained northeast along the front that stalled nearly parallel to the Balcones Escarpment, with 178 mm (7.02 in) of rain observed from 1405–1505 UTC along Onion Creek at Highway 183. A stronger downdraft and associated cold pool finally propagated these cells northeast after 1530 UTC (**Fig. 5.3e**) before they moved east of the study domain by 1925 UTC (**Fig. 5.3f**). The October event's heavier and more focused rainfall caused its Stage IV QPE to nearly double the May event's areal coverage (656 $km^2$ vs. 336 $km^2$) of rain totals over 152 mm (6 in), but the May event had more widespread (3040 $km^2$ vs. 2032 $km^2$) rain totals over 76 mm (3 in).

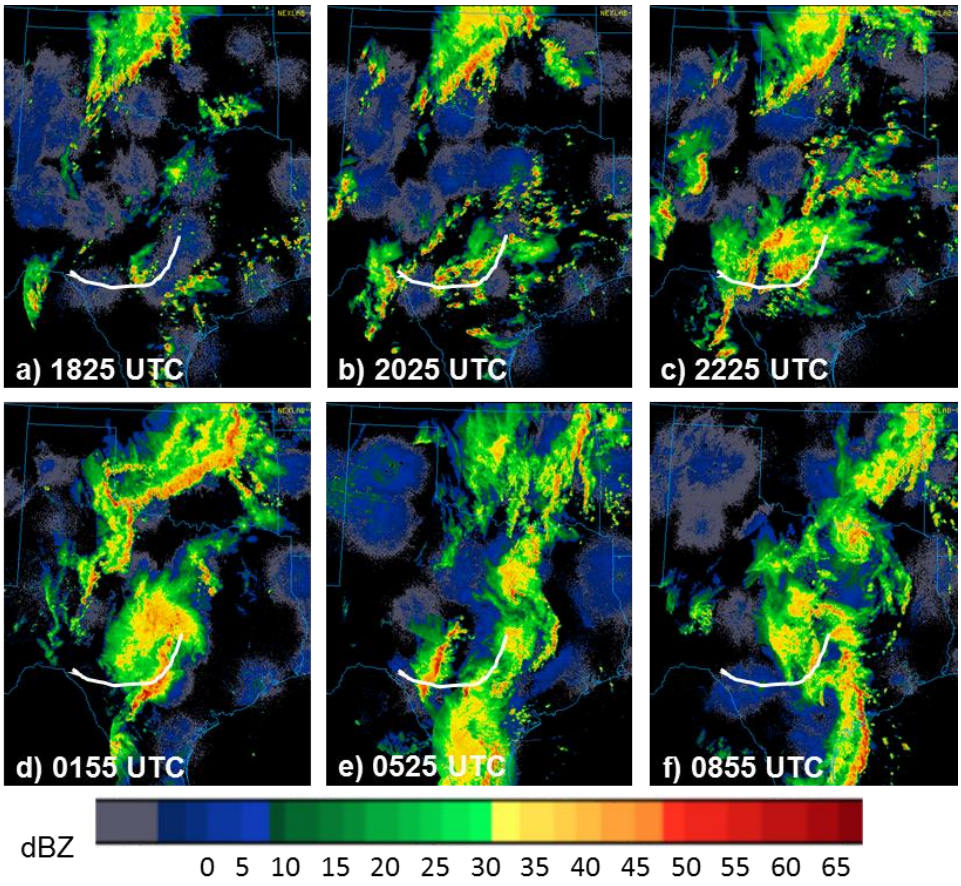


**Figure 5. 2.** Radar composite images at selected times from 1825 UTC 23 May 2015 to 0855 UTC 24 May 2015. The white line denotes the approximate location of the Balcones Escarpment. All images are obtained from an online archive maintained by the Mesoscale and Microscale Division of the National Center for Atmospheric Research (NCAR).

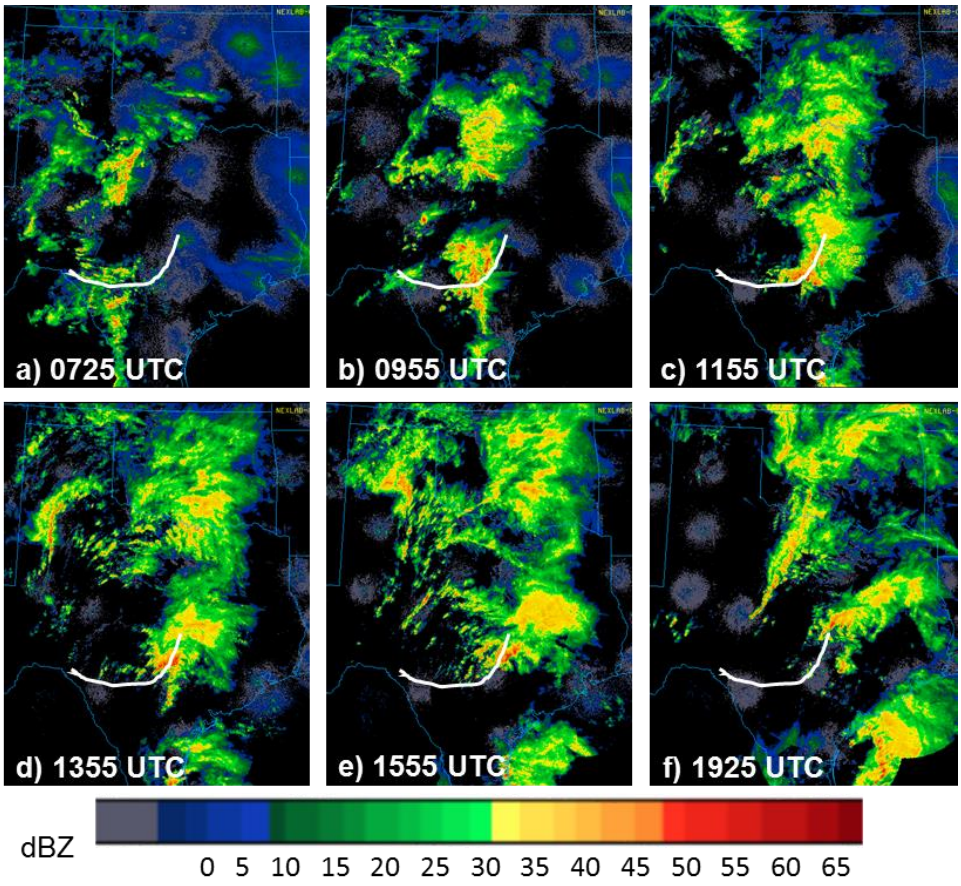


**Figure 5. 3.** Radar composite images at selected times from 0725 UTC to 1925 UTC 30 October 2015. The white line denotes the approximate location of the Balcones Escarpment. All images are obtained from an online archive maintained by the Mesoscale and Microscale Division of the National Center for Atmospheric Research (NCAR).

## 5.4. MODEL, DATA, AND EXPERIMENTS

### 5.4.1. Study domain

Our modeling domain covers south central Texas with an area of 91,571.3 km$^2$ (**Fig. 5.1a**, 27.96°N–31.34°N, 100.59°W–96.30°W). This domain is determined based on the intersected areas between the Hydrologic Unit Code-8 (HUC-8) watershed boundaries and 33 counties with the NWS Austin/San Antonio WFO, where two major cities (i.e. Austin and San Antonio) are located.

The domain incorporates 16,906 National Hydrography Dataset Plus (NHDPlus) river reaches and catchments, where a clear distinction in channel bed slope (**Fig. 5.1c**) and catchment slope (**Fig. 5.1d**) between the western and eastern sides of the Balcones Escarpment is displayed.

### 5.4.2. Model descriptions and augmentations

Under the modular architecture of WRF-Hydro, the Noah-MP LSM is tightly coupled to a vector-based river routing model (RAPID) (David et al. 2011) to simulate the flood discharges (hereafter WRF-Hydro-RAPID). Noah-MP is a state-of-the-art LSM originally developed for numerical weather predictions (NWPs) for vertical representation of hydrological processes. It is now used for physically-based spatially-distributed hydrologic simulations with the NWM (http://water.noaa.gov/about/nwm). RAPID is built upon the NHDPlus vector flowline network, which is the common geospatial framework adopted by the NWM. Because the NWM was in initial implementation as this project started and is under continuing development, we adopt the following strategies so that our results may be generalized within the conceptual framework of the NWM. First, the same Noah-MP parameterization options adopted by the NWM are used, including the free-drainage runoff option where surface runoff is calculated as excess water after precipitation supply infiltrates into soil layers; infiltration is solved using an approximate solution of the Richard's equation (Schaake et al. 1996) for four soil layers with thicknesses of 10 cm, 30 cm, 60 cm, and 100 cm. Subsurface runoff is calculated as the gravitational free-drainage from the bottom soil layer. Details of other hydrological processes can be found in Cai et al. (2014), Yang et al. (2011), and Zheng and Yang (2016). Second, we augment the hydrologic routing of the WRF-Hydro-RAPID model (see Appendix for details) by implementing (1) a time-variant overland flow delay function and (2) the Muskingum-Cunge channel routing using the same channel trapezoidal parameters obtained from the operational NWM.

| | This study | NWM |
|---|---|---|
| *Physical Model Chain:* | | |
| **Noah-MP LSM** | Same physical parameterization options (1-km resolution) | |
| **Overland Flow Routing** | Overland delay functions (Appendix) | Diffusive overland routing based on a 250-m grid |
| **Channel Routing** | Muskingum-Cunge routing based on NHDPlusV2; channel trapezoidal parameters obtained from http://water.noaa.gov/about/nwm | |
| *Other Modeling Capabilities:* | | |
| **Reservoirs** | Not represented | Simple level-pool routing with 1,260 major reservoirs in CONUS |
| **Data Assimilation** | No DA performed | Nudging-based DA with ~7000 gauges in CONUS |

**Table 5. 1**. Comparison between the model configuration of this study and the NWM. Only the historical/analysis configuration of the NWM is shown here; the short-, medium-, and long-range forecast configurations have some variations due to computational considerations.

**Table 5.1** provides a detailed comparison between the model configuration of this study and the NWM, which mainly differ in how reservoirs and overland routing processes are treated. Note our evaluation eliminates gauges downstream of reservoirs to minimize their effect on generalizations made about the NWM performance. In addition, the overland flow concentration time during major floods is often very short (< 100 s), and sensitivity tests with the delay function (not shown) suggest the flood hydrographs are minimally influenced in those cases. Therefore, the results here should have general implications to the NWM and other similar model configurations. In this study, the model results are not calibrated, meaning the parameters used are either directly linked to the land surface physical properties or estimated using widely documented empirical values in the literature; Manning's roughness coefficients are also obtained from the operational NWM. Although model calibration is a common practice for most hydrological studies, this study does not calibrate because of the increasing awareness of the "equifinality" problem suggesting

correct answers could be achieved for wrong reasons (Beven 2006). In addition, assessing a physically-based model with minimal calibration may be more informative in actual prediction scenarios, as they do not need to rely on historical data and the potentially incorrect assumption on hydrological stationarity (Milly et al. 2008) necessary for calibration. Although calibration is expected to benefit local gauges for engineering hydrology accuracy (Cho and Engel 2017), the goal of this study is to gain a process-based understanding of the model prediction skill at regional to local scales in an objective assessment.

### 5.4.3. Precipitation forcing: radar-based QPE products

The Stage IV (hereafter ST4) (Lin 2011) and Multi-Radar Multi-Sensor (MRMS) (Zhang et al. 2016) radar-based QPE datasets are used as the precipitation forcing inputs. ST4 has been the benchmark operational QPE over the past two decades, which has a 4-km and 1-h spatiotemporal resolution and is processed and quality controlled by hydrologists at 13 regional NWS River Forecast Centers (RFCs). In comparison, MRMS, which became operational in 2014, is automatically generated at a centralized location. One prominent feature of MRMS is its increased spatiotemporal resolution at 1 km and 2 min covering the CONUS. In this study, the benchmark ST4 QPE, the MRMS gauge-corrected (Q3GC), and the radar-only (Q3RAD) QPE products are utilized and evaluated due to their operational relevance and data availability.

### 5.4.4. Experimental design and model configuration

Three experiments are conducted for each event and are named after their precipitation forcing input datasets (see **Table 5.2**). All experiments are configured at 1-km resolution (in total 377×477 grid cells). The LSM and routing time steps are 1 h and 120 s, respectively. David et al. (2011) suggested a routing time step of 900 s for the stability criteria ($c \cdot \Delta t \leq \Delta L$) based on a median wave celerity (c) of 2.5 m $s^{-1}$ and a median reach length of 2.4 km, where c does not vary with time. This study uses, based on time-variant c with the Muskingum-Cunge, a much finer

routing time step (120 s) to allow for a maximum c of 20 m $s^{-1}$ for a median-length river reach, the stability of which was demonstrated for our simulations.

The two flood simulations start on 1 April 2015 and 1 September 2015 for the May and October events, respectively. The soil moisture initializations are interpolated from the 4-layer Noah LSM outputs of the North American Land Data Assimilation System-2 (NLDAS-2, 1/8°) (Xia et al. 2012). Although soil moisture in different LSMs may not be directly used from one model to another (Koster et al. 2009), we use this initialization strategy because Noah and Noah-MP are similar in many ways and a spin-up of over 50 days may help alleviate such problems. ST4 is used for the model spin-up period until two days before each event when different QPEs take place. Other forcing variables (radiation, temperature, humidity, wind, and pressure) are obtained from the NLDAS-2 forcing data.

| | **Experiments** | **Precipitation** | **Configuration** | **Spin-up** |
|---|---|---|---|---|
| **May 2015** | **ST4** | Stage IV QPE | 1-km and 1-hr for LSM; 120-s for routing | Initialized on 1 April and spin-up to 22 May |
| | **Q3GC** | MRMS gauge-corrected QPE | | |
| | **Q3RAD** | MRMS radar-only QPE | | |
| **October 2015** | **ST4** | Same as above | Same as above | Initialized on 1 September and spin-up to 28 Oct. |
| | **Q3GC** | | | |
| | **Q3RAD** | | | |

**Table 5. 2**. Experiments and model configuration of this study.

### 5.4.5. Reference in-situ measurements for evaluation

All best-available observational datasets are used to evaluate the model performance and obtain comprehensive quantitative assessment. These gauged datasets include: (1) daily precipitation reports from the Community Collaborative Rain Hail and Snow Network (CoCoRaHS) and the Cooperative Observer sites (COOP), (2) hourly precipitation intensity data from the NWS Hydrologic Automated Data System (HADS), (3) quality-controlled United States

Geological Survey (USGS) river discharge data, and (4) indirect USGS discharge measurements during gauge failures (see acknowledgements). CoCoRaHS is a dense rain gauge network collecting daily reports from trained volunteers (Reges et al. 2016). Although these datasets are subject to uncertainties due to the collection standards that may vary from user to user (Qi et al. 2016) and some gauged data may be ingested in the radar QPE products (Zhang et al. 2016), fully addressing this uncertainty is beyond the scope of this study and we have carefully used these datasets as independent references to satisfy our evaluation purposes.

In the study domain, 847 CoCoRaHS, 127 COOP, and 138 HADS gauges are utilized for precipitation evaluation (**Figs. 5.4a–5.4b**). 51 USGS Gage-II sites that are not downstream of reservoirs (out of the original 112 gauges; **Fig. 5.4c**) are used in discharge evaluation because the presented modeling framework does not account for reservoir operations. These 51 gauges are further separated into 27 forecast (i.e. gauges with existing NWS AHPS flood prediction capability) and 24 non-forecast points. USGS also separates the "referenced" and "non-referenced" gauge types, where 21 are classified as "referenced" (i.e. gauges in natural river basins) and 30 as "non-referenced" (i.e. gauges in basins with human alterations) to help delineate regional flood prediction skill in Section 5.5. Radar coverage map of the study domain is displayed in **Fig. 5.4d**.

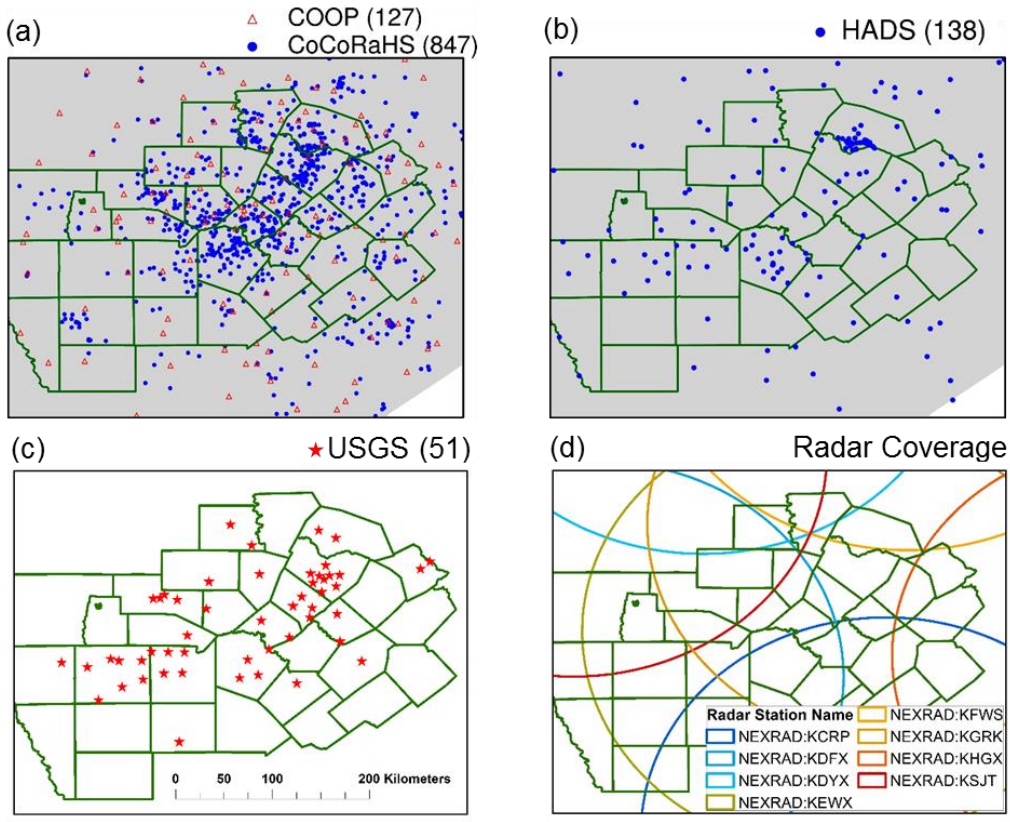


**Figure 5. 4.** Available in-situ measurements within the modeling domain. (a–c) show the gauges reporting 24-h precipitation, hourly precipitation, and discharge measurements, respectively, where the total number of available gauges is shown in the parenthesis. (d) shows the maximum radar coverage within the domain.

## 5.5. REGIONAL AND LOCAL FLOOD PREDICTION SKILL

### 5.5.1. QPE evaluation

#### *5.5.1.1. Event accumulated precipitation*

We first assess the event total precipitation against gauge observations. **Fig. 5.5** shows the scatter plot of accumulated precipitation between radar QPE and gauge observations, in which a pair of estimates (i.e. "R/G" pair) is accounted only when both of them are nonzero (Qi et al. 2016). To minimize the influence of CoCoRaHS gauge-reading times that may vary for a few hours around 1200 UTC, the scatter plots show a 24-h period (1200 UTC 23 May to 1200 UTC 24 May

2015) for the May event and a 48-h period (1200 UTC 29 October to 1200 UTC 31 October 2015) for the October event that capture rainfall for both events in their entirety. Statistics to quantitatively assess the QPE uncertainty include bias ratio (BR; sum of radar QPE divided by sum of gauged precipitation), correlation coefficient (CC), and root-mean-square-error (RMSE). BR values of 1 indicate a perfect match between the QPE and gauged data, whereas BR>1 (BR<1) suggests an overestimation (underestimation) by QPE.

All QPE sources exhibit reasonable accuracy in capturing both events' accumulated precipitation based on the range of their BR (0.86–1.11), CC (0.82–0.94), and RMSE (17.87–32.02 mm $d^{-1}$). MRMS consistently has lower estimates (lower BR) than ST4 for both events, but other statistics vary. ST4 has the best accuracy (**Fig. 5.5a**; BR closest to one and lowest RMSE) in the May event, whereas MRMS Q3GC (**Fig. 5.5b**) and Q3RAD (**Fig. 5.5c**) underestimate rainfall totals (BR<1). However, for the October event where ST4 (**Fig. 5.5d**) tends to overestimate, Q3GC (**Fig. 5.5e**) and Q3RAD (**Fig. 5.5f**) both display better accuracy by reducing the positive bias. MRMS reduces the RMSE by ~18 mm $d^{-1}$ (not shown) during the October event in areas where the heaviest rainfall occurred, which partly explained the better accuracy of MRMS during that event. Overall, the accuracy of Q3RAD (the radar-only MRMS product) is comparable to ST4 and Q3GC (the gauge-corrected MRMS product), which is promising considering Q3RAD does not require time for quality controls like ST4 and Q3GC (Cocks et al. 2016). However, Q3GC still demonstrates better accuracy than Q3RAD for the May event, suggesting that there are areas for the MRMS automatic algorithm to be improved to meet fine-scale accuracy needs.

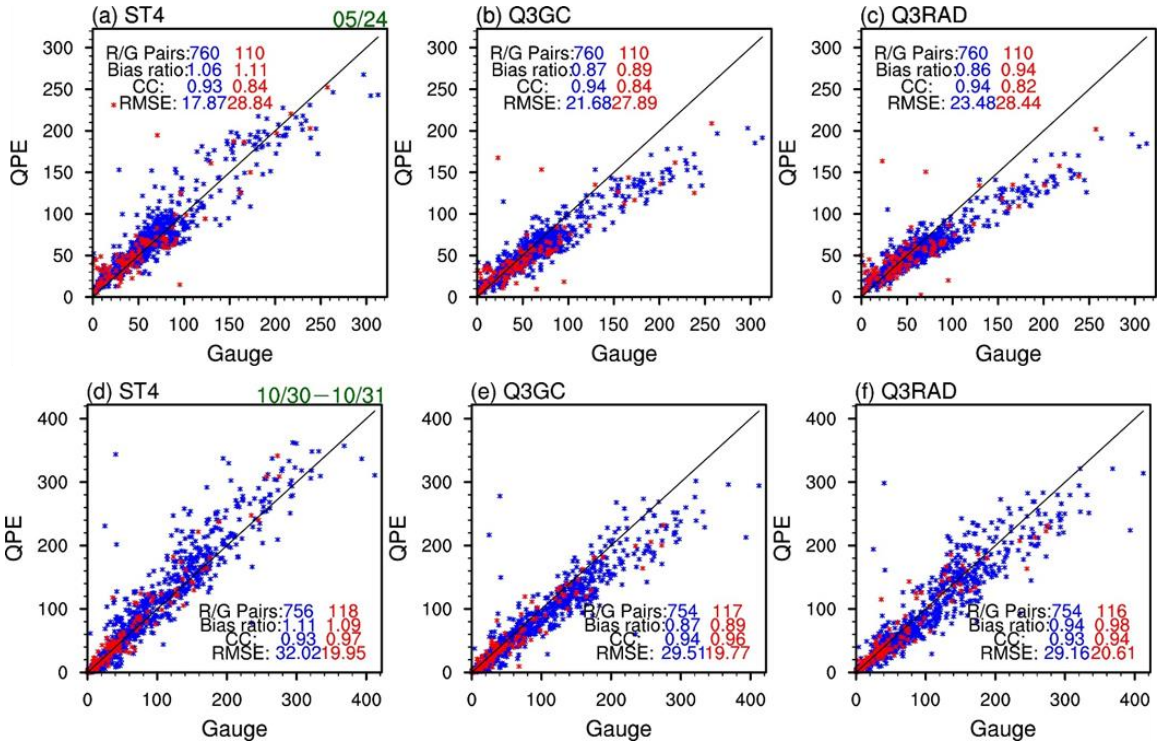


**Figure 5. 5.** Scatter plot for accumulated precipitation (mm) between radar QPEs (y-axis) and gauge observations (x-axis). Upper panels (a–c) show the 24-h accumulation for the May event, and lower panels (d–f) show the 48-h accumulation for the October event; the cycle starts and ends at 1200 UTC. Blue and red colors represent CoCoRaHS and COOP observations, respectively.

### *5.5.1.2. Maximum precipitation intensity*

The maximum precipitation intensity (MPI, mm $h^{-1}$) is assessed in **Figs. 5.6 and 5.7**. The May event has widespread MPI of 40–50 mm $h^{-1}$ for over ten counties, and all QPEs can reproduce the general patterns and magnitude for several localized heavy rainfalls. However, MRMS (particularly Q3RAD in **Fig. 5.6c**) fails to capture the most intense rainfall rates in excess of 60 mm $h^{-1}$. A scatter plot analysis (**Fig. 5.6e**) demonstrates hourly Q3RAD MPIs are no greater than 42 mm $h^{-1}$. To understand the May event's underestimation, we conducted a radar and statistical analyses using 2-min Q3RAD data (not shown) that suggests the problem is primarily caused by a truncation of rain rates at 1.78 mm per 2 min (53.4 mm $h^{-1}$) where the MRMS algorithm suspects there is hail contamination [i.e., "Maximum Estimated Size of Hail

(MESH)" greater than 0]. Readers may refer to Zhang et al. (2016) for more details about MRMS' algorithmic dependence on MESH and other parameters. Other possible causes for the underestimation may include (1) insufficient low altitude radar coverage to capture orographic enhancement and more efficient, tropical-like warm rain processes, and (2) the model analysis being too coarse to resolve locally moist tropical air masses [i.e., "Probability of Warm Rain" greater than 50–70%]. MRMS developers who assisted with our analysis are now implementing a new dual-pol QPE algorithm to reduce the MRMS algorithmic dependences on MESH (J. Zhang, personal communications), and we expect such improved QPE inputs to be key for more accurate river flow simulations in future studies. For the October event (**Fig. 5.7**), the rainfall was much more intense where large areas in Comal, Hays, and Travis Counties observed MPIs over 80 mm $h^{-1}$. Although all QPEs captured the major extreme rainfall band reasonably well, their accuracy still suffers in the lower rain rate range (< 50 mm $h^{-1}$, **Fig. 5.7e**) where QPEs tend to overestimate rainfall possibly related to false echoes. During this event, Q3GC (**Figs. 5.5e and 5.7b**) has the best accuracy in terms of both the event rainfall totals and the MPIs, which is translated into better regional flood prediction skill shown in section 5.4b that follows.

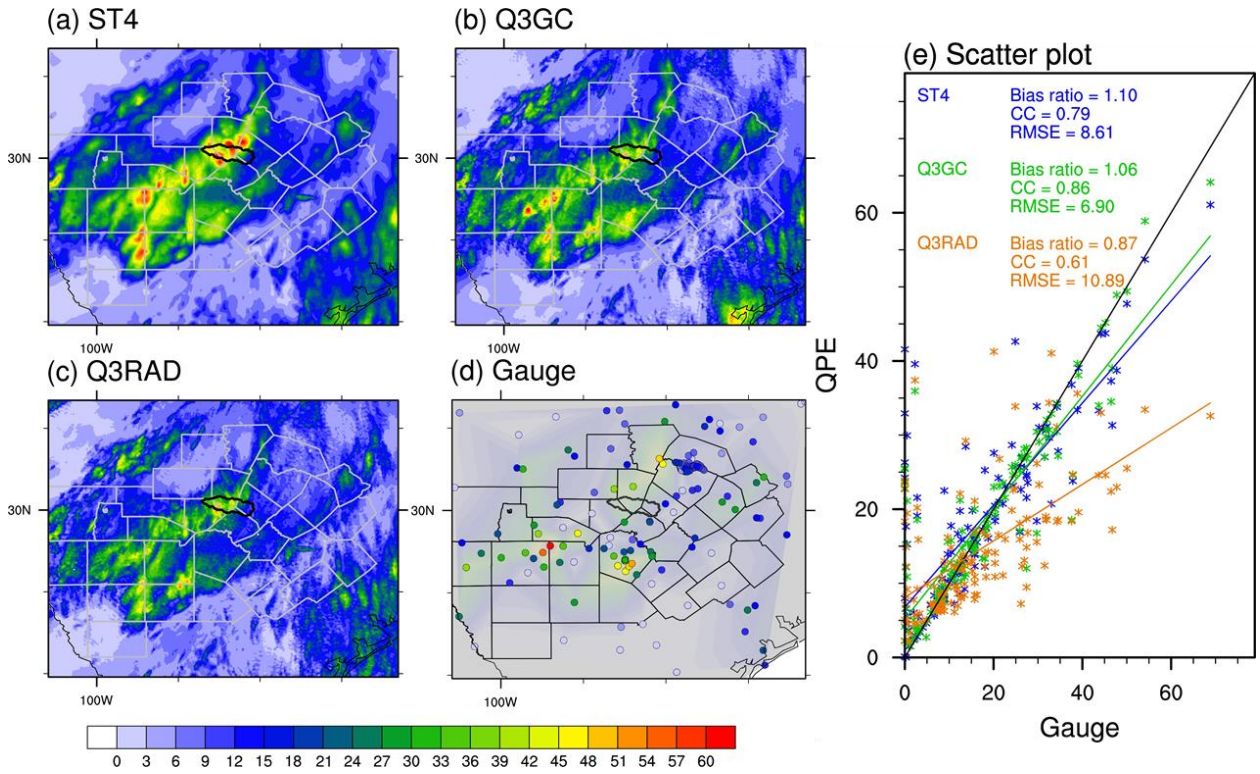


**Figure 5. 6.** The maximum precipitation intensity (MPI, mm $h^{-1}$) for the May event during 1200 UTC 23 May to 1200 UTC 24 May 2015. (a–d) show the spatial distribution of MPI in ST4, Q3GC, Q3RAD, and the HADS gauges, respectively. (e) shows the scatter plot along with regression lines and statistics displayed (black line indicates the 1:1 line).

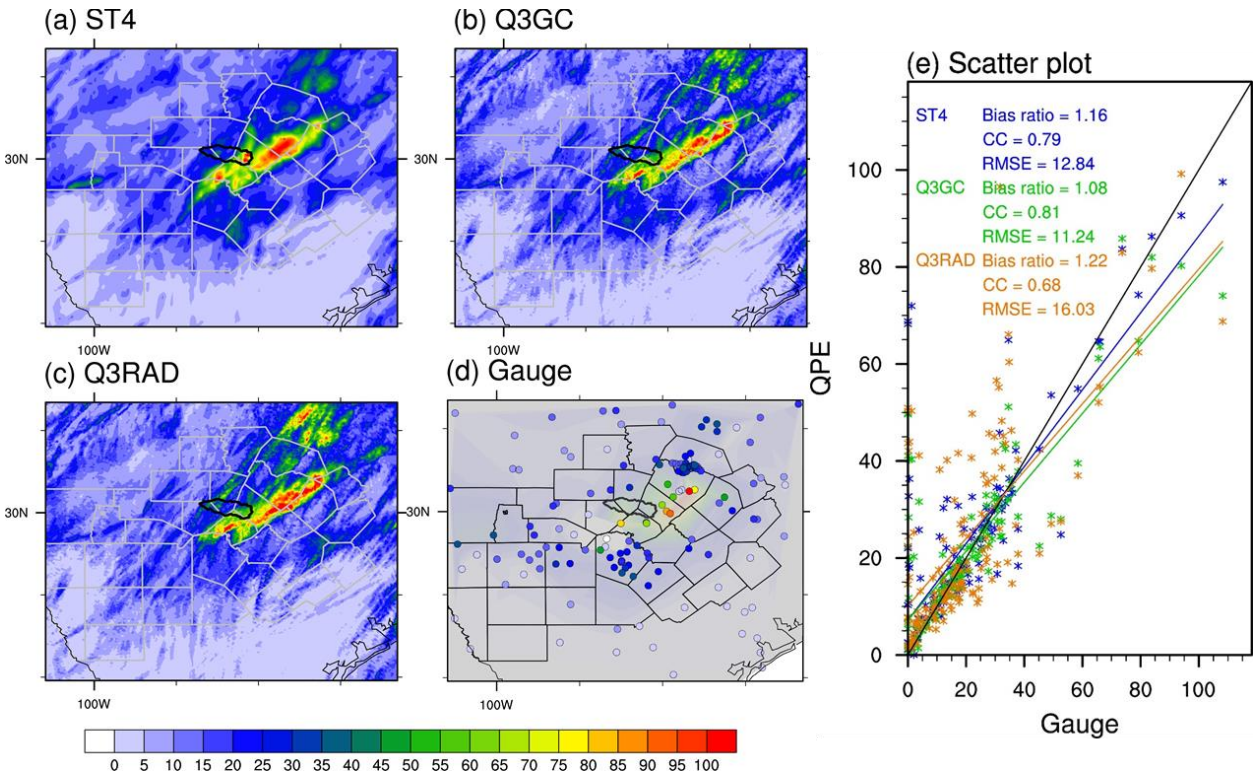


**Figure 5. 7.** Same as in Fig. 5.6, but for the October event during 0000 UTC 30 Oct to 2300 UTC 31 Oct 2015.

### 5.5.2. Regional flood prediction skill

In order to understand the flood prediction skill from a regional perspective seamlessly linking the river network flood wave propagation to local points that are key to emergency response, in this section we show a regional view of the flood prediction skill over south central Texas (**Fig. 5.8**) based on the Nash-Sutcliffe Efficiency (NSE) skill metrics (Gupta et al. 2009). NSE is the most commonly used quantitative measure of how well the model simulates the observed hydrograph. Regardless of which QPE is used in either event, the best flood predictive skill (based on NSE > 0.4 in **Fig. 5.8**) tends to be observed in regions receiving the most intense and greatest rainfall totals (cf. **Figs. 5.6 and 5.7**). This is very important to the strategic goals of the NWS operations, as the presented model clearly shows promise for flood prediction over the Hill Country where the most extreme rainfall and rapid river rises are observed. A graphical assessment of these simulated flood hydrographs is shown in Section 5.5.4.

However, model skill is not as good over regions that do not receive heavy rainfall for fast flood responses (see negative NSEs in **Fig. 5.8**). A further examination suggests the unsatisfactory discharge simulations are subject to different error sources. First, river discharge over regions with low rainfall totals can be dominated by baseflow. Four out of the 20 worst-performing gauges in the May event have consistently overestimated baseflow (not shown), which may point to the Noah-MP baseflow calculation problems needing future development and calibration. Second, other gauges with low NSEs almost always observe simulated flood hydrographs with more flashy and earlier peaks, which is a common problem also seen in the operational NWM (D.J. Gochis, personal communication). Muskingum-Cunge routing (hereafter "M-C") can be subject to 8–10% mass balance errors over flat areas (Todini 2007), which may partially explain the poor performance in parts of our domain. In addition, the earlier flood peaks indicate the model may have overestimated flood wave celerity c that may lead to an underestimated channel travel time k. In M-C, c is a function of the predicted discharge Q, the channel geometry, a factor β, and the Manning's roughness coefficient (see Appendix for M-C details). These factors are all subject to uncertainties because the presented model uses crude trapezoidal geometry assumptions and a simple but likely too large β factor. Moreover, no floodplains are represented in the model to buffer imminent flood waves. Because flow velocity and flood wave celerity are often the most poorly understood and least observed variables (McDonnell and Beven 2014), their physical parameterizations may be constrained in most hydrological models. Thus, reconsideration of the river network routing algorithm, better channel hydraulic geometry, a revised β factor, and representation of floodplains may be warranted to improve the flood prediction skill.

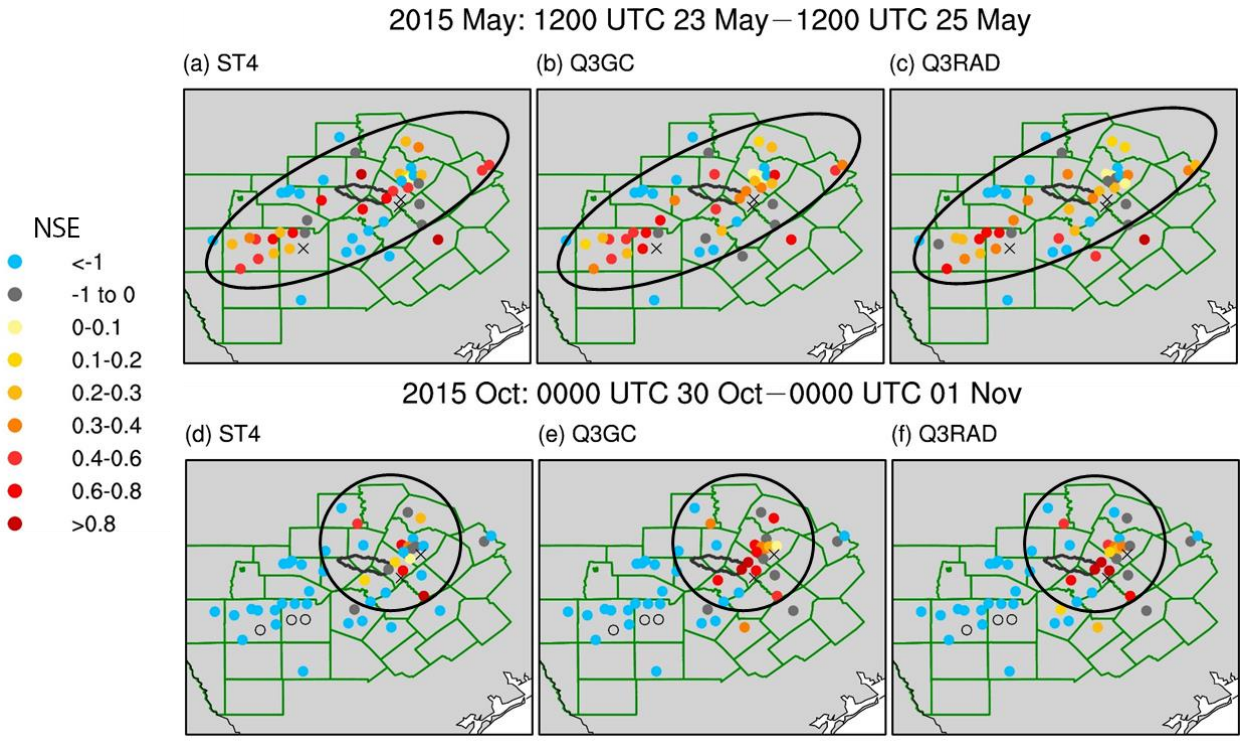


**Figure 5. 8.** Regional flood prediction skill for the May (upper panels, a–c) and the October (lower panels, d–f) events. The Nash-Sutcliffe Efficiency (NSE) at 51 gauges are shown. Crosses denote gauges with >60% missing measurements; hollow circles denote gauges with zero observed discharge. Black ellipses outline the approximate envelope areas for the storm cells.

### 5.5.3. Other factors influencing regional flood prediction skill

To better understand flood prediction skill at the regional scale, we summarize the NSE statistics in the boxplots using different criteria (**Fig. 5.9**). The same analyses are also conducted for the October event (**Fig. 5.17** in Supporting Information), but only the May event boxplots are presented for simplicity because the general conclusions are similar. **Fig. 5.9a** shows the regional median NSE is highest for ST4 and Q3GC (> 0.3), which agrees with the assessment that ST4 and Q3GC are equally good QPE inputs for the May event. Similarly, Q3GC and Q3RAD have the highest median NSE in the October event based on the best performing 20 gauges (**Fig. 5.17a**), matching the rainfall assessment that MRMS QPEs generally outperform ST4 in the Oct event. Interestingly, the skill interquartile range for MRMS is always narrower than that for ST4 in both events, indicating that compared to MRMS, ST4 has more variable flood prediction skill across

different sites. This range in flood prediction is reasonable because the rainfall inputs and LSM both have uncertainties across a large spatial domain, but variations in quality control performed by RFC hydrologists may help explain why ST4 struggles to generate more regionally-consistent skill relative to the automated MRMS product.

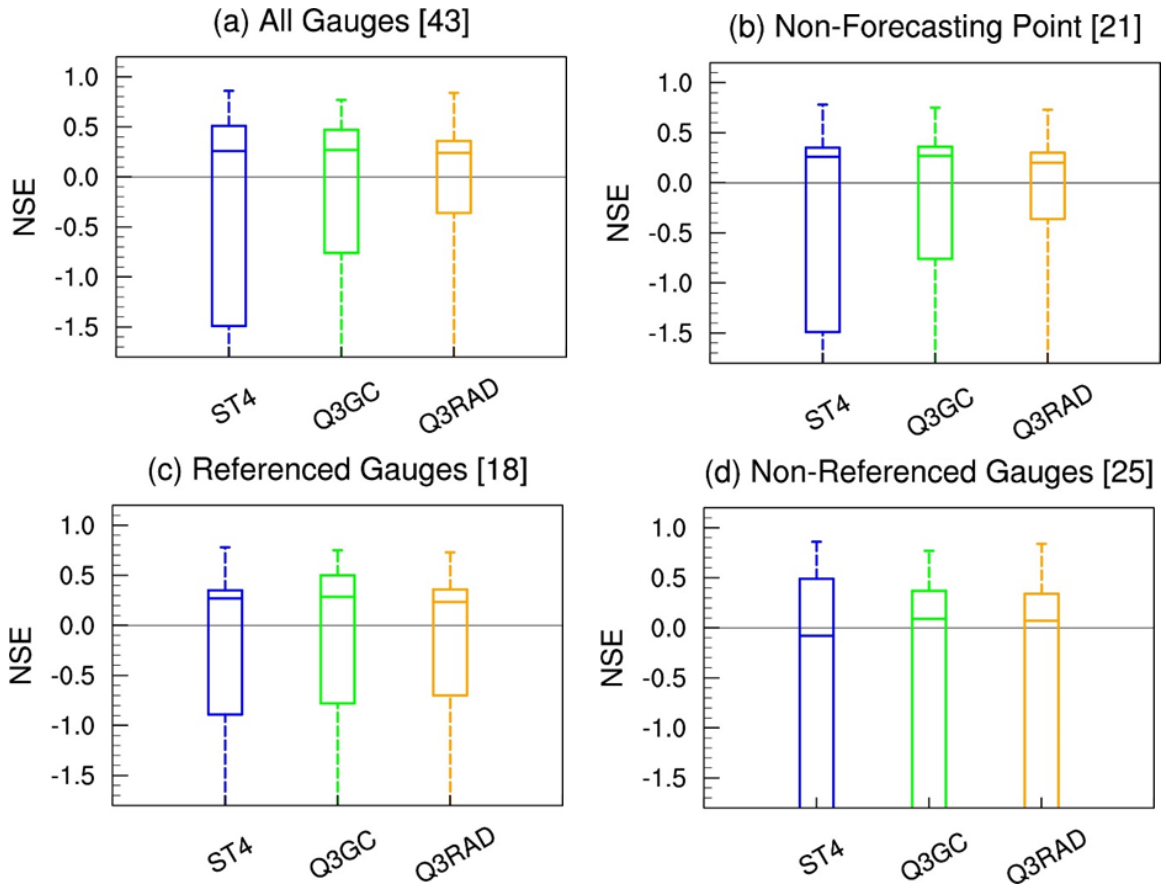


**Figure 5. 9.** Boxplot for regional flood prediction skill separated based on different gauge types. Whiskers show the maximum, 75$^{th}$, median, 25$^{th}$, and minimum NSE values for the May event. Gauges with >60% missing measurements and five gauges with consistent low NSEs (<–30) are excluded; the gauge number used to calculate the statistics is shown in the bracket.

The model skill for non-forecasting points is separately shown in **Fig. 5.9b**, which is comparable to that of all gauges (see similar median NSE in **Fig. 5.9a**). The NWM decreases the previous average modeling basin size from 1200 km$^2$ (Kitzmiller et al. 2013) to 2–3 km$^2$ now. The results here suggest the NWM can increase the forecast density by a factor of 700 (Lin et al. 2018)

while retaining reasonably good regional prediction skill. We also further separate the gauges into referenced (**Fig. 5.9c**) and non-referenced (**Fig. 5.9d**) gauges, where much lower skill for the latter group is seen. Non-referenced gauges mark watersheds with human-altered flows from water management and diversions. Although gauges downstream of reservoirs are excluded, the analyses point to the model's skillfulness in predicting natural river floods while suffering from other human alterations not yet accounted for in the model. The operational NWM incorporates a simple nudging-based data assimilation (DA) scheme to assimilate observed streamflow at ~7000 USGS gauges in its historical runs (**Table 5.1**), which might help alleviate the problem in **Fig. 5.9d** because human-altered flows could be reflected in the assimilated discharge observations. However, additional refinements of the DA capabilities and/or developments of modules to represent water management and human diversions may be warranted to further improve the regional model skill across a spatial domain.

**Figure 5.10** examines flood prediction skill by separating the gauges into groups with smaller (yellow) and larger (blue) drainage areas. Most prominently, MRMS QPEs demonstrate the most unique contributions to gauges with smaller drainage areas (especially those under 250 $km^2$). Note that the MRMS QPEs are generally less accurate than ST4 during the May event (cf. **Figs. 5.5–5.6**), but MRMS prominently outperforms ST4 for smaller drainage basin gauges (**Fig. 5.10**). This is most likely because ST4 cannot adequately capture the fine-scale rainfall patterns as limited by its coarser spatial resolution at 4 km (median NSE<0, blue boxplot in **Fig. 5.10d**). The same conclusion holds for the October event (see **Fig. 5.17b**) in which MRMS already exhibits better accuracy than ST4. In this event, the degree to which MRMS outperforms ST4 is also more obviously seen for gauges with drainage areas less than 320 $km^2$.

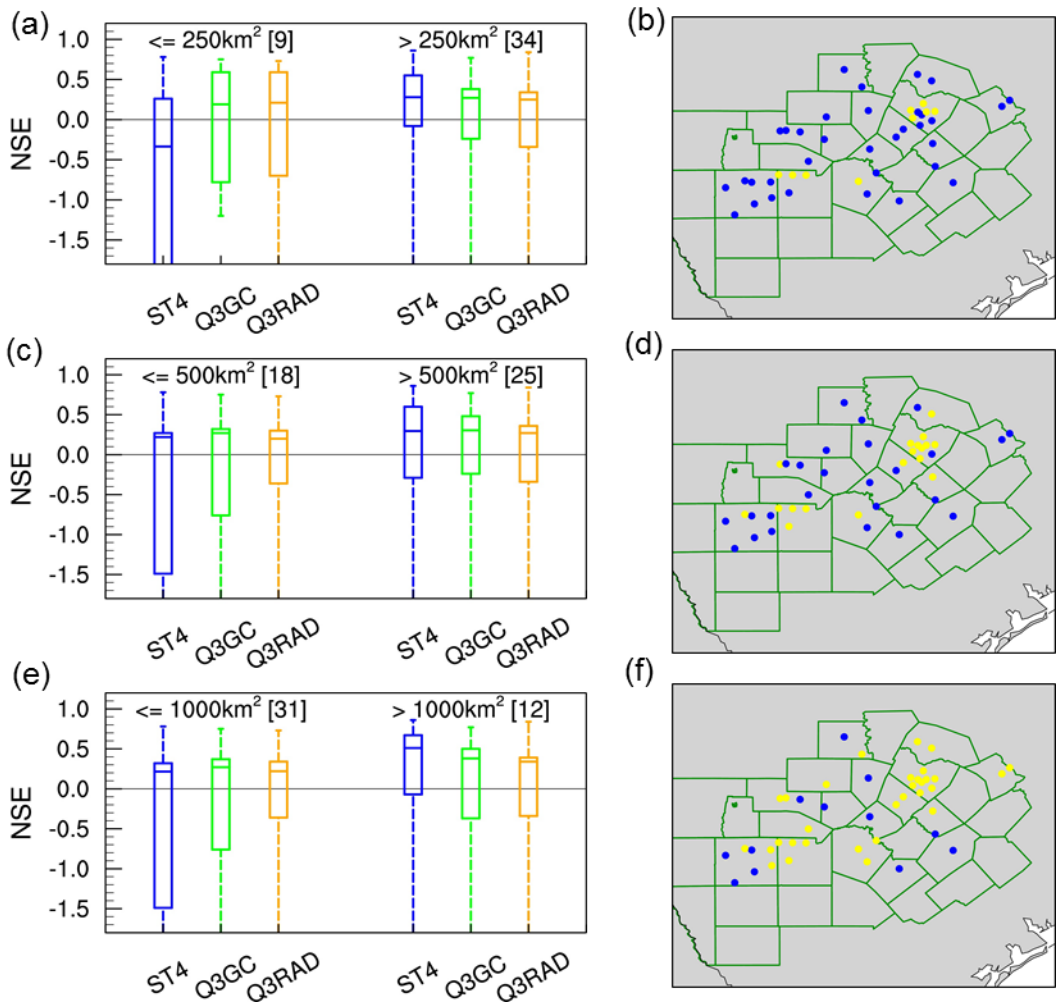


**Figure 5. 10.** Boxplot for regional flood prediction skill based on the drainage basin sizes. (a)(c)(e) show NSE boxplot and (b)(d)(f) show the gauge locations within the smaller- (yellow) and larger-area (blue) categories. Whiskers show the maximum, 75$^{th}$, median, 25$^{th}$, and minimum NSE values for the May event. Gauges with >60% missing measurements and five gauges with consistent low NSEs (<–30) are excluded; the gauge number to calculate the statistics is shown in bracket.

### 5.5.4. Flood hydrographs for selected local gauges of high strategic concerns

Flood hydrographs are shown in this section to provide a more direct assessment on the local-scale flood prediction skill. Six river gauges with high flood impacts and reasonable model skill are selected for both events. Consistent with our regional evaluation, precipitation forcing uncertainty serves as the dominant factor for the overall hydrograph and the peak magnitude prediction, matching the rainfall assessment that ST4 and Q3GC exhibit the best and comparable

accuracy in the May event, whereas MRMS outperforms ST4 for the October event. In addition, gauges with smaller drainage areas tend to benefit more from the MRMS QPEs, such as Hondo Creek (249.1 $km^2$), Shoal Creek (32.96 $km^2$), Slaughter Creek (22.72 $km^2$), and Walnut Creek (134.59 $km^2$) in **Fig. 5.12**. For both events, the predicted peak timing is ±2 h around the observed, and the rising limbs are captured better than the falling limbs. To issue effective flood warnings for emergency management decisions, accurate capturing of the peak timing and rising limbs is the topmost factor; thus, the results demonstrate good prediction skill even though their NSEs can be influenced by the imperfect falling limb simulations. For the May event, the predicted flood recessions are faster than observed, which may be a similar model behavior related to the "flashy hydrograph" problem with the current model parameterization (see discussions in Section 5.5.2). Both the regional and local evaluations suggest the channel routing may be oversensitive to changes in the flow velocity and wave celerity. This may have led to good model skill at gauges experiencing rapidly rising and falling floods, while showing unsatisfactory skill over regions without direct flood response; this problem will need to be revisited in future model development efforts.

Although the model performance displayed here is arguably imperfect, it exceeds the expected skill of current operational capabilities at the NWS Austin/San Antonio WFO, even though the model is not calibrated or tuned against certain events or precipitation forcing. Therefore, these results show promise for enhanced hydrological forecast capability with the NWM, in line with the goals of Braud et al. (2010) for a "model everywhere" that runs for gauged and ungauged basins. However, we also identified several areas that need to be revisited in future studies to improve regional flood prediction skill, which includes dealing with the precipitation uncertainties and spatial resolutions (Section 5.5.1), addressing the issue with the model flashiness over regions without an imminent flood response (Section 5.5.2), and incorporating modules or DA capabilities to represent human alterations on flows (Section 5.5.3).

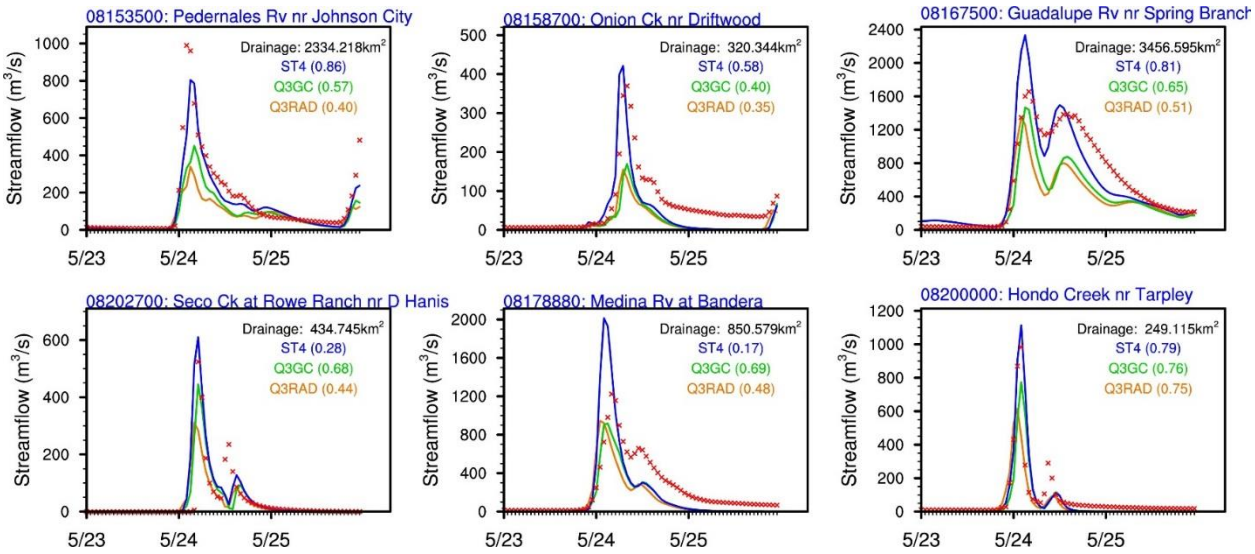


**Figure 5. 11.** Flood discharge hydrographs at six gauges of high strategic concerns in the May event. Red crosses stand for USGS observations and numbers in the bracket show the NSE values for each experiment (blue: ST4; green: Q3GC; orange: Q3RAD).

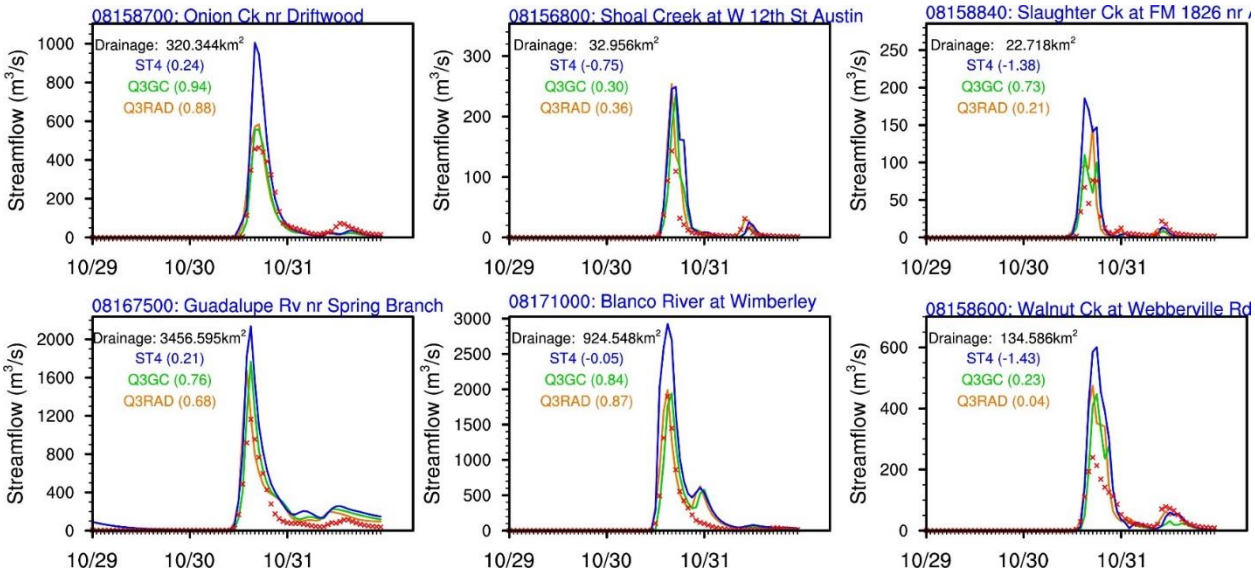


**Figure 5. 12.** Same as in Fig. 5.11 but for the October event. Red crosses stand for USGS observations and numbers in the bracket show the NSE values for each experiment (blue: ST4; green: Q3GC; orange: Q3RAD).

## 5.6. HYDROMETEOROLOGICAL DRIVERS FOR THE RECORD FLOOD FOR BLANCO RIVER AT WIMBERLEY

### 5.6.1. Contrasting flood responses for the May and October 2015 events

Considering the model performs well with NSE > 0.8 at some individual gauges, the best-simulated model results can be used to obtain a better mechanistic understanding of the drivers and flood responses at specific locations for the May and October events. In this section, we analyze both events at the Blanco River at Wimberley (USGS #08171000; WMBT2) whose peak flood stage in the May event (13.7 m) exceeded the previous record by 3.5 m. Due to similarities in meteorological characteristics and antecedent soil moisture in both events, the May record flood at WMBT2 that exceeded the October crest by 5.6 m was difficult to resolve. Therefore, an analysis to understand the hydrometeorological drivers of both events at WMBT2 is needed.

**Figure 5.13** shows that the two events observed similar event total precipitation, initial surface soil moisture, and average surface runoff in the drainage basin of WMBT2, yet the spatial distribution is concentrated in different locations. During the May event, most of the heaviest rainfall (> 250 mm) fell on the basin's southwestern part, whereas the October event had slightly higher rainfall totals (> 320 mm) close to the basin outlet (**Fig. 5.13a**). Antecedent soil moisture conditions were very saturated prior to both events (> 0.3 $m^3$ $m^{-3}$; Fig. 13b) – one after the wettest April and May for Texas (Furl et al. 2018) and one after receiving ~150 mm of rainfall associated with the remnants of Hurricane Patricia on 23–24 October 2015 (Rogers et al. 2017). The rainfall combined with the nearly saturated soil resulted in slightly higher surface runoff during the October event closer to the basin outlet (**Fig. 5.13c**) relative to the May event.

Interestingly, the observed flood in May had a return period well in excess of 100 years (USGS PeakFQ Flood Frequency Analysis) (Veilleux et al. 2014), whereas the October event only reached a 10-year return period level (see red dots in **Fig. 5.14**). The rising limb of the May event was also much steeper, where the stage increased by nearly 9 m over a 2-h interval (0330 to 0530 UTC on 24 May 2015) to surpass the record. However, the best simulated modeling results forced

by ST4 (NSE = 0.68) for the May event and Q3RAD (NSE = 0.87) for the October event were only able to produce slightly different flood responses, where the peak stage is better simulated for the October event (see blue lines in **Fig. 5.14**). It is therefore interesting to examine what factors may have led to the observed contrasting flood responses, and why the model struggles to capture the full magnitude of the record flood for the May event.

**Fig. 5.15** displays the WMBT2 drainage basin physiographic features and the events' rainfall locations, which offers a qualitative explanation for why the October flood response was less prominent than that in May. The upstream headwater catchments over which most of the May event's rain fell feature steeper channel bed slopes (> 0.01 in **Fig. 5.15**). These steeper catchments together with the relatively straight upstream Blanco River in the western portions of the basin could propagate large and fast flood waves to WMBT2, along with the eastward propagating rainfall. In comparison, the October event mainly operated on the less steep and more sinuous downstream portions of the Blanco River, which may serve to naturally suppress big flood waves. A closer look at the simulation results (blue line in **Fig. 5.14a**) suggest the model is able to capture these differences to some degree, with a steeper and higher flood peak simulated for the May event exceeding the 25-yr return period. This is primarily because the major land physiographic features (e.g. channel and catchment slope) are incorporated in the model's physical parameterization. However, there is still a significant underestimation of the peak stage for the May record flood, and a quantitative understanding could be facilitated through a series of model sensitivity tests with the best observational constraints in Section 5.6.2.

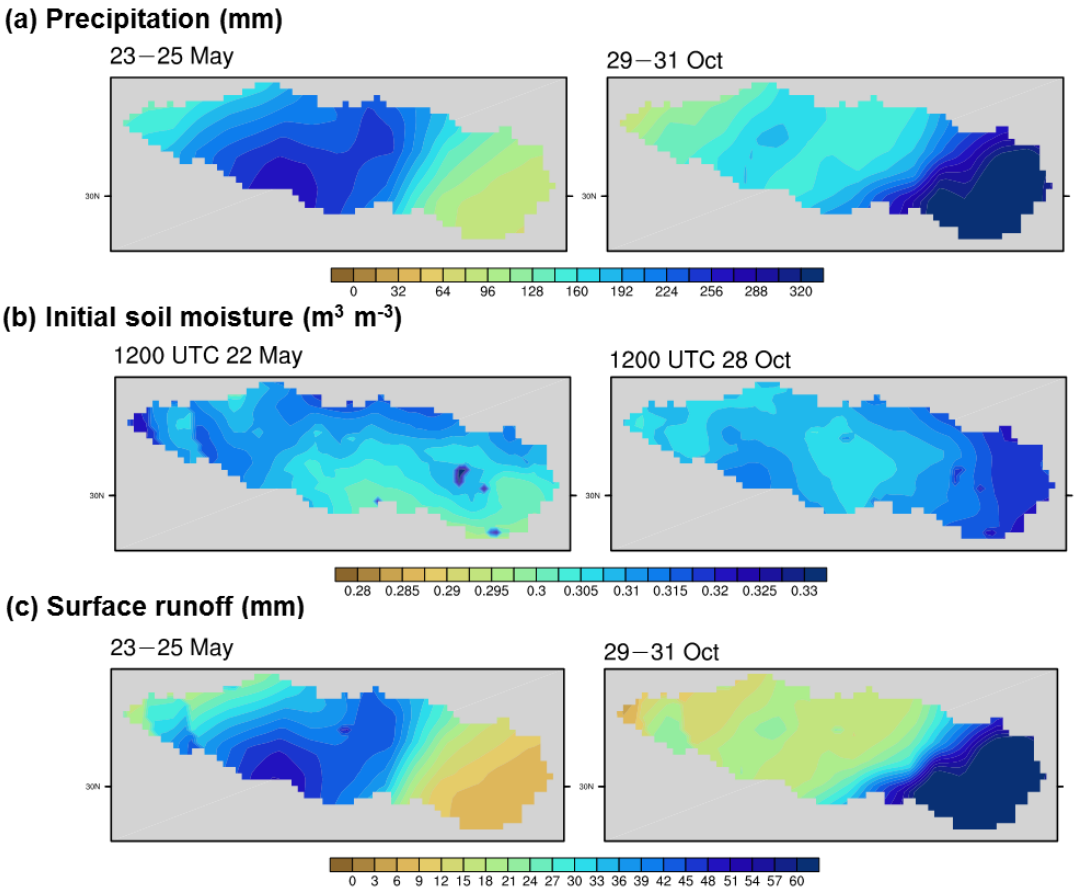


**Figure 5. 13.** Hydrometeorogical conditions for the drainage basin of WMBT2 during the May and October floods (ST4 for May; Q3RAD for Oct). (a) shows the event total precipitation (mm); (b) shows the initial soil moisture conditions ($m^3$ $m^{-3}$); and (c) shows the event-averaged surface runoff (mm). The cycle starts and ends at 1200 UTC of the first and the last day, respectively.

### 5.6.2. Sensitivity Analyses for Quantitative Understanding of the Local Record Flood

Several error sources may cause the underestimation of the peak stage height. First, the underestimation may be associated with errors in precipitation forcing, the Noah-MP soil moisture calculation, and the flow velocity calculation with the routing model. Second, measurement errors may also occur with the indirect discharge estimations and the rating curve utilized. Third, there might be errors in neglecting flowing debris objects as they may obstruct flow at narrower portions of the river channel and increase viscosity within the river channel itself. Unfortunately, these processes are not understood well enough to be included in our current modeling framework.

Complex surface–subsurface interactions with the existence of karstic spring features in south central Texas (Looper and Vieux 2012) may be another factor currently not accounted for. However, based on limited observations at a main karstic spring on the Blanco River above WMBT2 (not shown), the spring contributions may play a minor role as they are several orders of magnitude lower than that from surface runoff during big flood events.

Below, precipitation forcing, soil moisture, and flow velocity are examined individually in an attempt to shed light on how they play a role in reproducing the record flood peak in May 2015.

- The most accurate ST4 precipitation for this event is 14.6% higher than gauged rainfall based on 15 CoCoRaHS and COOP rain gauges in the WMBT2 basin (not shown). This suggests QPE uncertainties alone cannot explain the underpredicted peak stage height.
- For soil moisture, the model predicts a ~80% basin-average degree of saturation for the surface soil layer (see **Fig. 5.13b**, based on the dominant soil type of clay loam and maximum soil moisture of 0.465 $m^3 m^{-3}$). Because the basin has an overall soil depth of 30–60 cm, sensitivity tests are conducted by saturating the top two soil layers (sat1: 0–10 cm, sat12: 10 cm–40 cm) in the model's initialization based on each grid cell's soil type. It is found that sat1 and sat12 increase the peak stage height by 0.1 m and 1.4 m, respectively, the latter of which at least exceeds the 1929 record flood (**Fig. 5.16**). These sensitivity tests agree with previous studies, suggesting the degree of saturation for deeper soils, compared to the surface soil layer, may play a more important role in the streamflow and peak stage prediction (Rajib et al. 2016) due to their larger water storage.
- For flow velocity, the only reference data are indirect estimations obtained from the Guadalupe–Blanco River Authority (GBRA) and some post-event measurements from the USGS (see acknowledgements). According to the May event flow animation (not shown here), the flood wave propagating down from the Blanco River upstream of its confluence with the Little Blanco River lagged behind the latter's flood wave by ~1 h. In order to test what the flood responses would be if the two flood waves arrived at the confluence of the Blanco River and the Little Blanco River simultaneously, we manually increased the flow velocity by

multiplying the model-predicted velocity by a factor of 1.2 for the first and second Strahler order streams above the conjunction (see **Fig. 5.18** for the Order1 and Order2_Blanco river reaches). Aside from making the two waves join simultaneously in this sensitivity experiment (sat12v12), all other conditions are the same as sat12. The peak stage height for sat12v12 is further increased by another 0.9 m above sat12, demonstrating the flow velocity spatial combination may also have played an important role in producing the record flood peak. The sat12v12 sensitivity run predicts the flood wave crest 1–2 h earlier than observed, which may be explained by debris flows on the sinuous Blanco River that might have temporarily slowed down the flood wave. Nevertheless, increasing the flood peak stage by simply modifying the flow velocity for lower order streams just 20 km upstream suggests accurately representing flow velocity, particularly at the headwaters (McDonnell and Beven 2014), is indeed critical for predicting the May 2015 record flood peak.

**Figure 5.16** also shows sat12v12 generally has the closest-to-observed flow velocities at two time points – one around the peak timing and one during the recession stage (blue arrows). This further confirms the crucial role of initial wetness in deeper soils and flow velocity at headwater catchments, as tested by sat12v12, in producing skillful flood peak prediction. The May 2015 record flood at WMBT2 by nature involves more complex spatiotemporal variations in the network's velocity field than the October event, posing more challenges for the model to predict the peak stage height well. Our results indicate a certain spatial combination of the velocity field in the river network may have contributed to the catastrophic record flood in May 2015, and hydrological models need to represent the spatial combination well enough to be able to predict such a flood. Studies evaluating future flood events in the WMBT2 basin will benefit from four additional river gauge measurements that have been installed since May 2015, which could improve our mechanistic understanding of the precipitation drivers, flood responses, and the model prediction skill in the future.

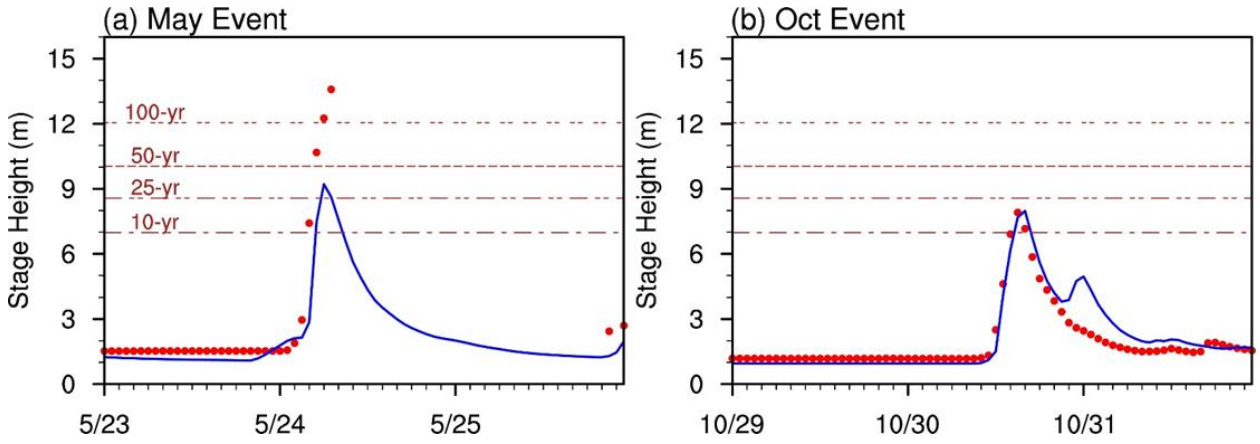


**Figure 5. 14.** Flood stage height at WMBT2 during the (a) May and (b) October events in UTC time. Red dots denote observations and blue lines denote the best-simulated results (ST4 for May; Q3RAD for Oct). Discharge is converted to stage height using the USGS rating curve at WMBT2. Flood return period (dashed lines) is determined using the USGS PeakFQ flood frequency analysis with the annual maximum streamflow excluding the two examined events.

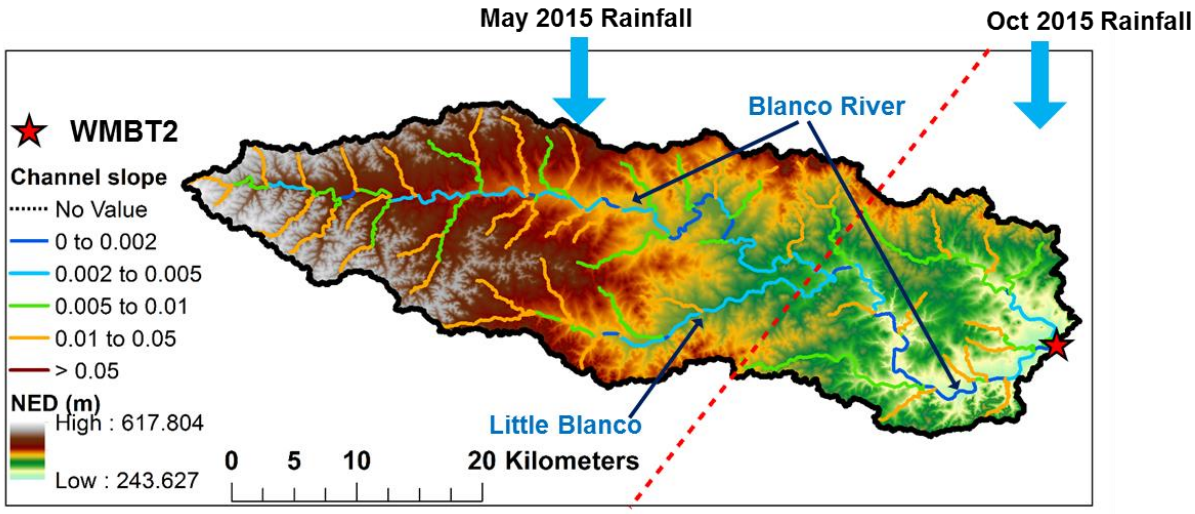


**Figure 5. 15.** The landscape physiographic features (topography and channel bed slope) in the drainage basin of WMBT2. Red dashed line separates the approximate rainfall locations for the May and October 2015 events.

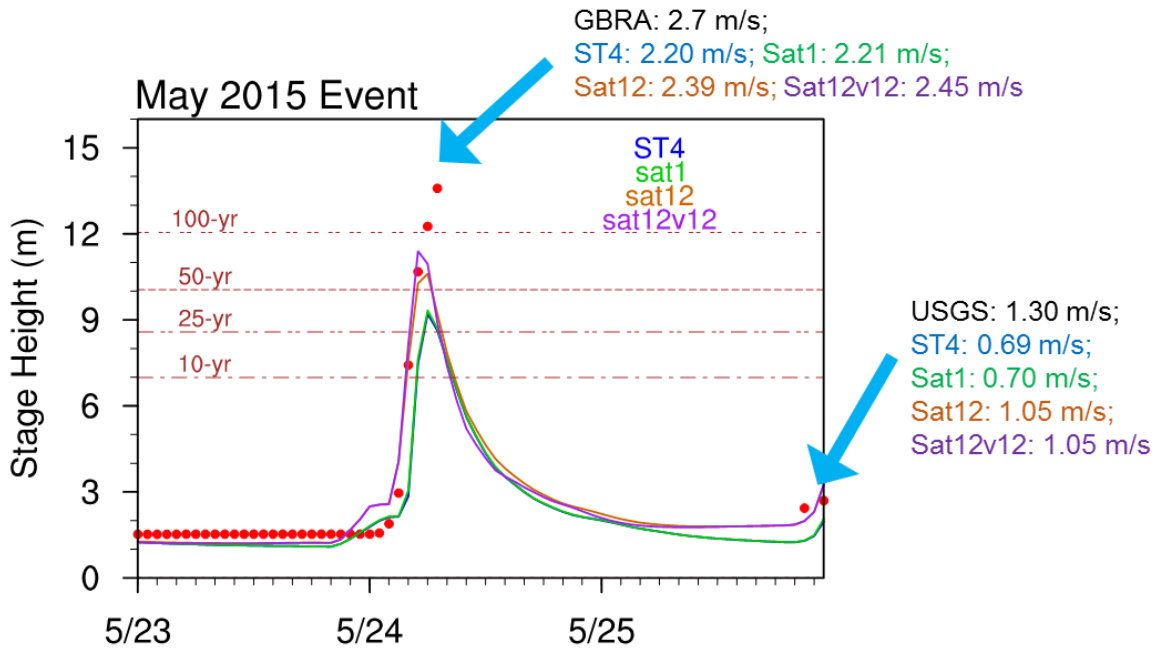


**Figure 5. 16.** Sensitivity tests for the May 2015 flood simulation at WMBT2. Red dots denote observations, and color lines denote results from different experiments. Blue arrows point to two time points where the model-predicted flow velocities are compared with those estimated by the GBRA and the USGS.

## 5.7. CONCLUSIONS AND DISCUSSIONS

This study aims to evaluate how a continental-scale water dynamics model linking meteorology, land surface hydrology, and river channel routing performs in the Texas Hill Country, a well-known flooding hotspot of the United States. Towards this goal, we employ a physically-based model configured to resemble the NWM to provide necessary information that may aid NWS forecasters in flood emergency response. Based on two high-impact flood events in the Hill Country during May and October 2015, comprehensive assessments are conducted using the best observational constraints (i.e. 947 daily rain gauges, 138 hourly rain gauges, and 51 USGS river gauges) to identify the model's strengths and weaknesses. Major conclusions of the regional-scale model evaluation are summarized below.

- Using two radar-based QPE products (i.e., ST4 and MRMS), we found the minimally calibrated flood prediction skill generally follows the event precipitation accuracy from the evaluation at a range of gauges. This highlights the critical role of precipitation forcing inputs in accurate flood prediction at the regional scale.
- Due to the increased spatial resolution, MRMS clearly outperforms ST4 particularly in gauged basins with small drainage areas with room left for improvement for its automated algorithm.
- Overall, the model not only shows good performance at gauges experiencing rapid rises of floodwater from causative rainfall, but also performs reasonably well for gauges that currently are not NWS AHPS forecast points. These results show great promise for enhanced hydrologic forecasting capability for emergency response.
- However, the model suffers from poor prediction skill over areas without imminent flood response and those with human-altered flows, suggesting the need for revised channel routing algorithms and explicit considerations of human alterations.

Note that the current NWM implements a simple level-pool routing for 1,260 major man-made reservoirs and initial DA capabilities, which may help alleviate problems with human-altered flows. However, the results pointing to the channel routing are applicable to the NWM and similar model configurations using the same Muskingum-Cunge routing and channel parameters. In future studies, more advanced channel routing considering floodplain representations and reservoir operations, combined with refined DA algorithms, and terrain-based routing under continuing development of the NWM need to be explicitly tested in order to quantify how the inclusion of these modules could improve the regional model performance.

In addition to model evaluation, this study also probes into the hydrometeorological factors contributing to the contrasting flood responses at WMBT2 in May and October 2015. This investigation is to facilitate mechanistic understanding on the basin's unexpected flood response in May 2015, and further dig into reasons on why the model struggles with predicting the record flood. Conclusions from this local assessment are listed below.

- Using the best-simulated results, it is shown the location of the intense rainfall in combination with the land physiographic features are the key factors that may help explain the two events' different flood responses.
- By conducting a series of model sensitivity tests, we demonstrate the record flood peak in May 2015 can be better obtained by tuning the deeper soil initial wetness and the flow velocity field in the river network. This clearly highlights the crucial role of subsurface soil saturation and the network's flow velocity in producing an unexpected flood response at WMBT2.

The quantitative tests presented in this study not only offer insights into the causes for the local record flood in May 2015, but also help point to possible reasons on why the model struggles to predict the flood peak. Other random factors that are not incorporated in the current modeling framework as well as the NWM may also play a role, such as subsurface contributions from karstic springs and debris flows that may pile up flood water at certain locations along the river. Due to a lack of observations, however, our understanding and modeling capabilities for these processes are still limited, and thus their targeted observations are warranted in the future.

## 5.8. SUPPORTING FIGURES

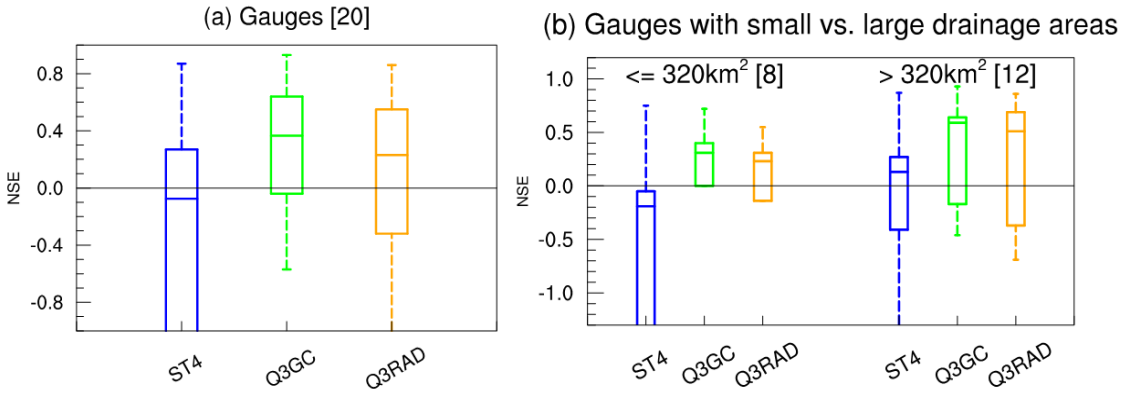


**Figure 5. 17.** Regional simulation statistics for the Oct 2015 event. (a) shows the NSE boxplot for the top 20 ranking gauges, similar to Fig. 7; and (b) shows the statistics of these 20 gauges as separated by the 320 $km^2$ threshold, similar to Fig. 5.8.

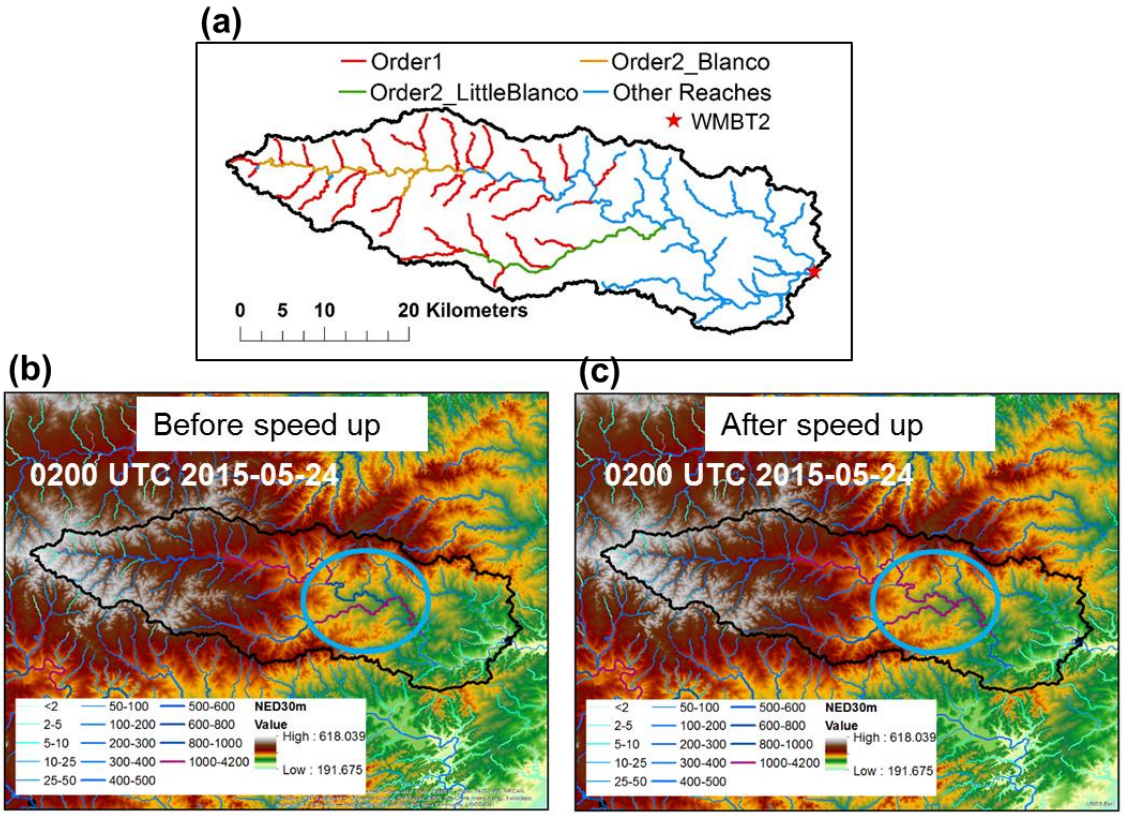


**Figure 5. 18.** Sensitivity experiment for sat12v12. (a) shows river reaches in the WMBT2 basin, (b)(c) show the snapshot at 0200 UTC 24 May 2015: two flood waves before and after the speeding up for Order1 and Order2_Blanco river reaches by a factor of 1.2.

## 5.9. APPENDIX: MODEL AUGMENTATION DETAILS

### 5.9.1. Time-variant overland flow delay

At each LSM time step i, surface and subsurface runoff generated on Noah-MP grid cells are averaged based on intersected areas with NHDPlus catchments. This produces lateral inflows ($Q^{i}_{lateral}$, in $m^3\ s^{-1}$) for the "blue-line" river reaches, which first travels along NHDPlus catchments as overland flow, and then is routed through river channels. Based on the geospatial attributes of the NHDPlus, overland flow travel time ($T_{ov}$) can be computed using Eqn. (5.1) (Cho and Engel 2017; Neitsch et al. 2011), in which $T_{ov}$ is a function of the catchment flow length ($L_{cat}$), average

catchment slope (slp), catchments' roughness coefficient ($n_{cat}$), and lateral inflows ($Q^i_{lateral}$). Eqn. (5.1) thus represents a time-variant form of $T_{ov}$ that varies with $Q^i_{lateral}$, where higher $Q^i_{lateral}$ can lead to smaller $T_{ov}$. In the equation, $n_{cat}$ is the roughness coefficient related to land cover types obtained from the WRF-Hydro look-up table (Gochis et al. 2015); $L_{cat}$ (m) is approximated as half of the catchment area divided by channel length, similar to Li et al. (2013)); and slp (m m$^{-1}$) is derived from terrain analysis in ArcGIS (Fig. 5.1c).

$$T_{ov} = \frac{L_{cat} \cdot {n_{cat}}^{0.6}}{slp^{0.3} \cdot {Q^i_{lateral}}^{0.4}} \quad \text{Eqn. (5.1)}$$

Taking into account of the $T_{ov}$ delay, the lateral inflow that is able to contribute to the vector flowlines ($Q^i_{lateral,r}$) is computed using Eqns. (5.2–5.3) (Neitsch et al. 2011), where $S^{i-1}$ and $S^i$ are runoff stored in the catchment at the previous and current routing time step, respectively ($S^0$=0); $\alpha$ is the runoff lag coefficient, or the characteristic overland flow time, which is calculated for each catchment using the $T_{ov}$ equation of Neitsch et al. (2011) assuming an average net incoming flux of 6.35 mm hr$^{-1}$; and $f$ is an adjustment factor currently set as 0.5.

$$Q^i_{lateral,r} \cdot \Delta t = (Q^i_{lateral} \cdot \Delta t + S^{i-1}) \cdot (1 - e^{-\frac{f \cdot \alpha}{T_{ov}}}) \quad \text{Eqn. (5.2)}$$

$$S^i = Q^i_{lateral} \cdot \Delta t + S^{i-1} - Q^i_{lateral,r} \cdot \Delta t \quad \text{Eqn. (5.3)}$$

### 5.9.2. Time-variant channel flow: augmentation of the Muskingum-Cunge routing

After $Q^i_{lateral,r}$ reaches the NHDPlus flowlines, a Muskingum-Cunge channel routing algorithm (hereafter M-C) is implemented to augment the original RAPID Muskingum routing for two reasons. First, M-C is adopted by the operational NWM and results from M-C may help inform strengths and weaknesses of the current NWM. Second, M-C is a more physically-based method than Muskingum because it uses the channel geometry characteristics and the predicted discharge to update its routing parameters k and x at each time step; thus, it may be more adapted to actual

prediction applications where no observed data could be obtained prior to a particular event as required by the Muskingum method.

M-C and Muskingum solve the same Eqn. (5.4) to predict river reach j's outflow at time point i ($Q_j^i$), where $C_1$, $C_2$, $C_3$, and D are coefficients computed using k and x; and $C_4$ is a coefficient for lateral inflow contributions (Eqns. 5.5–5.8).

$$Q_j^i = C_1 Q_{j-1}^{i-1} + C_2 Q_{j-1}^i + C_3 Q_j^{i-1} + C_4 \quad \text{Eqn. (5.4)}$$

$$C_1 = (k \cdot x + \Delta t/2)/D \quad \text{Eqn. (5.5)}$$

$$C_2 = (\Delta t/2 - k \cdot x)/D \quad \text{Eqn. (5.6)}$$

$$C_3 = [k(1-x) - \Delta t/2]/D \quad \text{Eqn. (5.7)}$$

$$C_4 = Q_{lateral,r}^i \cdot \Delta t/D \quad \text{Eqn. (5.8)}$$

$$D = k(1-x) + \Delta t/2 \quad \text{Eqn. (5.9)}$$

In Muskingum, k and x are constant in time. In M-C, k is updated at each time step as channel length ($\Delta L$) divided by instantaneous wave celerity ($c$); x is updated as a function of peak flow ($Q_p$), top width ($tw$), channel bed slope ($bs$), $\Delta L$ and $c$ (Eqns. 5.10–5.11) (Bedient and Huber 1988), the parameter expressions of which make M-C a diffusion wave model.

$$k = \Delta L/c \quad \text{Eqn. (5.10)}$$

$$x = 0.5 - Q_p/(2 \cdot tw \cdot bs \cdot c \cdot \Delta L) \quad \text{Eqn. (5.11)}$$

Using Manning's equation, instantaneous wave celerity $c$ is estimated using average flow velocity $v$ as a function of channel roughness ($n_{chan}$), hydraulic radius ($hr$), and $bs$ (Eqn. 12); and $\beta$, a coefficient to relate $c$ with $v$. $\beta$ equals to 5/3 for cross sections where the hydraulic radius can be approximated by the depth, where a more accurate $\beta$ should decrease after exceeding the channel capacity, and increase again with deep flow in the floodplain (Merkel 2002). For preliminary use in the initial NWM and this study, a constant $\beta$ of 5/3 is used but future studies need to better represent this factor based on more detailed considerations for the river channel and floodplain hydraulics. At each routing time step, $hr$ is first updated using the Newton–Raphson

iterative method that converges when the steady-state $Q_j^{i-1}$ are satisfied for the trapezoidal channel shape, after which all parameters are updated using Eqns. (5.5–5.13).

$$v = 1/n_{chan} \cdot hr^{2/3} \cdot bs^{1/2} \quad \text{Eqn. (5.12)}$$

$$c = \beta \cdot v \quad \text{Eqn. (5.13)}$$

When augmenting the RAPID model with the M-C routing algorithm, vector-based calculations using the Portable, Extensible Toolkit for Scientific Computation (PETSc; Balay et al. 2001) are retained because they contribute to the high computational efficiency of RAPID.

## 5.10. ACKNOWLEDGEMENT

This study is supported by the UCAR COMET NWS Partners Project (Award No. Z16-23463). David R. Maidment and Xing Zheng from UT-Austin are thanked for offering project suggestions. Jian Zhang and the MRMS group is acknowledged for their helpful discussions. Special thanks go to David J. Gochis (NCAR), Youcun Qi (NOAA), Arezoo Rafieei Nasab (NCAR), Alec MacDonald (USGS) for helping with the NWM, MRMS, HADS, and river discharge/velocity data during gauge failure. Other datasets are downloaded from publicly available sites, including ST4 (https://data.eol.ucar.edu/dataset/21.093), MRMS QPE (http://mtarchive.geol.iastate.edu/2015/), Global Historical Climatology Network gauged precipitation (ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/), and USGS quality-controlled discharge (https://waterdata.usgs.gov/nwis). The model codes and simulations for this study are available upon request to the corresponding author. The authors declare no conflicts of interest.

# Chapter 6: Summary and Future Perspectives

To achieve a better understanding, modeling, and predicting capability of the coupled terrestrial hydrologic cycle from the land perspective, this dissertation focuses on (1) introducing enriched geospatial information to advance the snow and river routing component of two state-of-the-art LSMs (CLM4 and Noah-MP), (2) assessing their roles in the temperature, precipitation, and streamflow prediction, and (3) adapting them for actionable uses.

## 6.1. DISSERTATION SUMMARY

In Chapter 2, I focus on assessing the role of snow DA in seasonal temperature prediction. I found an improvement of 5%–25% locally over regions with prominent snow DA updates such as the Tibetan Plateau (TP), Siberia, and northern North America, where more pronounced and longer-lasting improvements are seen in GRACE DA than MODIS DA. Interestingly, we found the higher latitudes observe delayed improvements while the lower latitudes observe immediate improvement. This spatiotemporal pattern could be explained by the incoming solar radiation that is key to establishing the regional snow–atmosphere coupling.

In Chapter 3, I focus on assessing the potential contribution of snow DA in the Asian monsoon seasonal forecast based on the findings in Chapter 2 and the knowledge that the TP and Siberian snow is key to the Asian monsoon predictability. I found that among all the Asian monsoon subcomponents, the robust improvement signal for precipitation forecast is observed in central north India, a region receiving the most Indian summer monsoon rainfall with great agricultural implications. We clarify several complexities needed for communicating snow DA to improved rainfall forecast, which includes the regional snow-atmosphere coupling strengths, the DA uncertainties, and the monsoon's sensitivity to thermal effects over the Eurasian landmass.

In Chapters 4 and 5, I present the model coupling work that implements a vector-based routing model in the community WRF-Hydro framework. To inform the flood emergency response with the newly operational United States National Water Model (NWM) that shares a similar model configuration, I further extend the work by focusing on assessing the model performance

over the Texas Hill Country, a well-known flooding "hotspot" of the CONUS. We use all the best available observational constraints to systematically evaluate the model's flood prediction skill during two high-impact flood events. In addition, we use best modeling results and model sensitivity experiments to gain hydrometeorological insights into a local flood. The results point to three reasons that can help explain the relatively unexpected record-breaking flood at Blanco River at Wimberley in May 2015. They include location of the intense rainfall and the basin physiographic features, the initial wetness of the subsurface soil layers, and the spatial combination of the flow velocity field in the network upstream of the location of interests, all of which have challenged the successful prediction of the record event's flood peak. The results call for the need for an integrated meteorological–hydrologic–hydraulic modeling framework.

## 6.2. APPLICATIONS EXTENDED FROM MY PHD RESEARCH

As stated in Chapter 1, LSMs have now been widely used beyond its traditional realms of weather and climate prediction (Chapters 2 and 3), along with recent advancements in the LSM hydrological parameterizations as well as increased model resolutions (Chapters 4 and 5). However, to truly adapt the "top-down" LSM modeling approaches to address direct societal concerns that emerged in fine-scale hydrologic applications, there are still several missing pieces.

During my Ph.D. studies, I also worked on demonstrating the feasibility of regional-scale flood inundation mapping, a challenging task with traditional hydrologic/hydraulic approaches (Merwade et al. 2008), by connecting the large-scale hydrologic modeling framework (Chapters 4 and 5) with a terrain-based method. The preliminary result (**Fig. 6.1**) was obtained by working with Dr. David R. Maidment's research group in the University of Texas at Austin, where a GIS-based method for terrain analysis was recently developed to characterize channel geometry for all 2.7 million NHDPlus flowlines across the CONUS (Liu et al. 2017; Zhang et al. 2018, *in revision*). This method can be directly coupled to large-scale hydrologic modeling frameworks for regional-

scale flood inundation mapping at low computational cost, where flood inundation areas are estimated by integrating weather patterns and major hydrologic processes with terrain information.

In the Summer Institute of the second National Flood Interoperability Experiment (NFIE) (Maidment 2017), weather forecasters, first responders, and students were gathered in the National Water Center (NWC) in Tuscaloosa, AL, to participate in a mock flood emergency response exercise. The goal is to reimagine future flood forecasting practices that seamlessly unifies the meteorological/hydrologic/hydraulics modeling approaches in a generic framework. To prepare the streamflow simulation and inundation mapping data needed for this exercise, I specifically set up the model domain for the Tuscaloosa County, which has 2,118 NHDPlus river reaches and an area of ~4000 $km^2$ using similar model configurations as presented in Chapter 4 and the radar-based StageIV data as the rainfall inputs. The final inundation mapping results were then displayed using the ArcGIS image service (acknowledgement to Deng Ding, ESRI), which allowed all participants to zoom in and out for their interested neighborhoods. As pointed out by the first responders in this emergency response exercise, a few local buildings that were inundated in the flood (circled in **Fig. 6.1b**) were captured by the workflow presented above, which was promising given the model is uncalibrated and suitable for large-scale applications with low computational needs. The demonstration then motivated research projects to compare this simple terrain-based approach with the 2-D hydrodynamic model outputs and that from remote sensing images (Zhang et al. 2018), where an accuracy of 56% was suggested, with potential improvements up to 67% if catchments that were not flooded locally but received flood water from their adjacent catchments were considered separately. Although this approach arguably integrates all sources of uncertainties and is an approximate method, it seamlessly links synoptic weather, LSMs, river routing, and inundation mapping in an earth system modeling approach (Sivapalan 2018). In our view, it is an emerging research theme to bridge the gap between traditional hydrologic/hydraulics modeling that attempts to "scale-up", and the atmospheric/LSM modeling that attempts to "scale-down" towards its "hyperresolution" goal (Wood et al. 2011). However, we also note that future studies are warranted to understand and carefully assess the uncertainties cascading from weather to

hydraulics (similar to materials presented in Chapter 5), such that similar modeling frameworks can be truly adapted for use in addressing the societal needs.

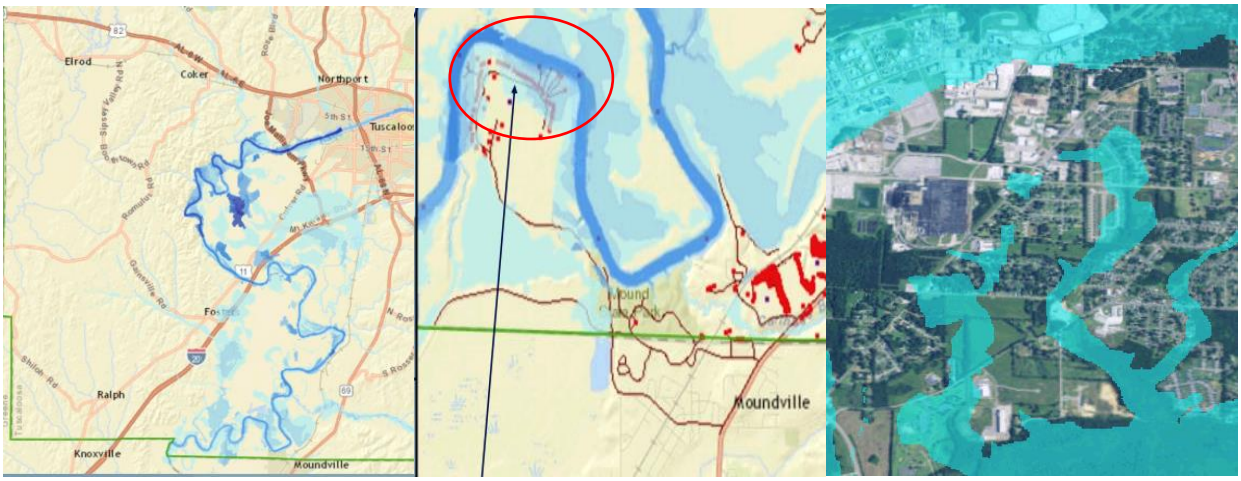


**Figure 6. 1.** Regional flood inundation mapping for Tuscaloosa County, Alabama, USA, during the Boxing Day flood in 2015. (a) shows the inundated area calculated for the 2,118 NHDPlus catchments, (b) shows the inundation overlaid by building address points, and (c) shows a neighborhood-scale flood inundation on top of the Google Images. Discharge simulation is performed using the uncalibrated WRF-Hydro-RAPID (Chapter 4). Discharge is converted to stage height using the synthetic rating curves derived from the Height Above Nearest Drainage (HAND) method from Liu et al. (2017) and Zheng et al. (2018, *in revision*).

## 6.3. CAVEATS AND FUTURE WORK

Although I have demonstrated a few promising hydroclimate applications that can benefit from enriched geospatial information to the LSMs, which contributes to our understanding on the practical predictability of the terrestrial hydrological cycle. However, several important caveats remain to be further studied beyond the materials presented in this dissertation. Below, I am listing a few major ones that need more research investigations.

- **Chapters 2 and 3: Seasonal climate prediction**

**First,** one main limitation with Chapters 2 & 3 is the limited data samples in our statistical analyses, which constrained a better quantitative understanding on how much DA respectively

contributes to the *anomaly* and *bias* component of a prediction. For a seasonal climate forecast to be useful, e.g., providing accurate forcing inputs to seasonal streamflow prediction for water resource applications, it is usually much more important and difficult to get the anomaly component correct rather than the bias component. This is because the systematic bias in a forecast is arguably more easily corrected using post-processing techniques, assuming we know the bias of a model. Although both Chapter 2 and Chapter 3 attempted to understand the DA contribution to the anomaly component by using $r^2$, scatter plots, and several reference datasets, however, possible sampling noise might still exist to interfere with the true signal due to the limited snow DA data samples (7 years). Therefore, we have not clearly separated the contribution of DA to each of these two components. To this end, future studies using longer periods of land DA products in a coordinated multi-model experiments might be warranted to better understand this problem.

**Second,** our investigation to quantify the role of DA has not focused on anomalous snow years, during which time more pronounced contributions from snow DA might be exhibited. For example, the relationship between the extreme snowstorm over the TP in 2003 and the ensuing wetter-than-normal East Asian monsoon is of great interests to researchers (Seol and Hong 2009) to understanding monsoon anomalies associated with anomalous antecedent snow conditions. Thus, investigating the role of DA in extreme snow conditions is also warranted.

**Third,** to better inform actual seasonal climate prediction that accounts for the dynamical interactions among all major components of the climate system, results in these chapters need to be extended using fully-coupled modeling experiments that turn on the interactions between the ocean, atmosphere, land, and sea ice modules.

**Fourth,** carefully designed snow DA experiments are also needed to better understand which regions are more important for the Asian monsoon forecast (Senan et al. 2016; Turner and Slingo 2011) and the mechanisms responsible for translating DA to detectable signals in rainfall forecast. These mechanistic studies might help with identifying regions with true signals versus regions where noises may more likely to occur.

- **Chapters 4 and 5: Large-scale flood prediction**

For Chapters 4 & 5, I also identified several future research areas of immediate importance. **First**, there is the need to better represent the river channel and floodplain hydraulics in the current routing algorithm. It is a common practice for engineering-based models (Brunner 1995) to use detailed channel cross sections collected in the field for flood prediction purposes. While "at-a-station" hydraulic geometry (Singh and Zhang 2008) is considered in this approach, it is highly data demanding and not easy to be "scaled up", raising the question as to whether such amount of information is necessary (Gleason and Smith 2014) for flow discharge predictions at the synoptic scale. The study presented in Chapter 5 uses the crude trapezoidal channel geometry that is related to the Strahler stream orders, which is a widely-adopted practice for large-scale hydrologic modeling (e.g. Neitsch et al. 2011). However, the introduction of the NHDPlus fabric to traditional LSM modeling has offered new opportunities to take advantage of new data attributes developed by the GIS community, potentially providing the geospatial information needed to bridge the gap between these two modeling approaches. An interesting next step would be to use the newly estimated channel geometry data "lumped" at the reach scale at the 2.7 million river CONUS reaches (Liu et al. 2017) to improve the model physical realism, and assess how much information is needed for the synoptic-scale flow prediction. In line with this effort, it is also believed that more hydraulics-based routing algorithms need to be revisited when feeding the model with better channel geometry representations.

**Second,** the advent of satellite altimetry data that characterize the water depth and channel widths (Durand et al. 2010) can also potentially offer new insights into the large-scale streamflow dynamics from space. In the future, I am interested in developing an integrated method merging both remote sensing-based and terrain-based approaches in better characterizing channel geometry, understanding fluvial geomorphology, and improving the predictive capability of the flow dynamics in similar large-scale modeling frameworks.

**Third,** as mentioned in Chapter 1, I have used radar-based rainfall inputs in Chapters 4 and 5 because these two chapters focus more on the model's performance in simulating the hydrologic

responses. To make actual flood predictions at real time, however, weather forecasts generated from coupled weather/climate models (e.g. WRF) are usually fed into the hydrologic modeling framework to make such predictions. This source of uncertainty, however, is highly likely to be the most important one and it may overwhelm other uncertainties as examined in Chapters 4 and 5 (e.g. model resolution, coupling interface, initial soil moisture estimation, and flow velocity representations in the model). Thus, huge challenges remain for streamflow forecasting of any type, which are not easily resolvable in the foreseeable future. In fact, according to the experiences of the NWS forecasters at the Austin/San Antonio Weather Forecasting Office (WFO), the operational NWM that takes in the real-time weather forecasts is still not producing satisfactory flood forecasts as good as what is presented in Chapter 5, and the reason may be ascribed to the limited weather forecast accuracy especially at the storm scale. However, we expect the results in these two chapters to be useful because they demonstrate challenges other than weather forecasts (e.g. initial wetness conditions in deeper soil layers and the network flow velocity representations) that are still constraining the flood forecasting skill. They also improve our understanding on the complex nature of floods, especially those ones exceeding historical records, upon which future research using a joint view combining weather, land hydrological processes, and river hydraulics is warranted to enhance our existing flood forecasting capability.

Overall, both the atmospheric and streamflow prediction parts of this dissertation lead me to a common future research theme to use information theory to boost the understanding of enriched geospatial information in physically-based models. Some recent studies (DelSole and Shukla 2010; Kumar et al. 2018) were conducted along these lines to understand the relationship between the model biases and the seasonal climate prediction skill, as well as the information content of different satellite remote sensing datasets. These areas seem central to improving the understanding of this dissertation's research topics, which will warrant more investigations.

# Bibliography


Arnault, J., S. Wagner, T. Rummler, B. Fersch, J. Bliefernicht, S. Andresen, and H. Kunstmann, 2016: Role of runoff–infiltration partitioning and resolved overland flow on land–atmosphere feedbacks: A case study with the WRF-Hydro coupled modeling system for west africa. *Journal of Hydrometeorology*, **17,** 1489-1516.

Ashley, S. T., and W. S. Ashley, 2008: Flood Fatalities in the United States. *Journal of Applied Meteorology and Climatology*, **47,** 805-818.

Balmaseda, M., and D. Anderson, 2009: Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, **36**.

Bedient, P. B., and W. C. Huber, 1988: Hydrology and floodplain analysis.

Best, M. J., and Coauthors, 2015: The Plumbing of Land Surface Models: Benchmarking Model Performance. *Journal of Hydrometeorology*, **16,** 1425-1442.

Beven, K., 2006: A manifesto for the equifinality thesis. *Journal of hydrology*, **320,** 18-36.

Bierkens, M. F., and Coauthors, 2015: Hyper-resolution global hydrological modelling: What is next? Everywhere and locally relevant. *Hydrol Process*, **29,** 310-320.

Blanford, H. F., 1884: II. On the connexion of the Himalaya snowfall with dry winds and seasons of drought in India. *Proceedings of the Royal Society of London*, **37,** 3-22.

Bonan, G. B., and S. Levis, 2006: Evaluating aspects of the community land and atmosphere models (CLM3 and CAM3) using a dynamic global vegetation model. *Journal of Climate*, **19,** 2290-2301.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 571-657.

Braud, I., H. Roux, S. Anquetin, M.-M. Maubourguet, C. Manus, P. Viallet, and D. Dartus, 2010: The use of distributed hydrological models for the Gard 2002 flash flood event: Analysis of associated hydrological processes. *Journal of Hydrology*, **394,** 162-181.

Brown, R., and B. Brasnett, 2010: Canadian Meteorological Centre (CMC) daily snow depth analysis data. *Environment Canada*.

Brunner, G. W., 1995: HEC-RAS River Analysis System. Hydraulic User's Manual. Version 1.0.

Caesar, J., L. Alexander, and R. Vose, 2006: Large–scale changes in observed daily maximum and minimum temperatures: Creation and analysis of a new gridded data set. *Journal of Geophysical Research: Atmospheres*, **111**.

Cai, X., Z.-L. Yang, C. H. David, G.-Y. Niu, and M. Rodell, 2014: Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. *Journal of Geophysical Research: Atmospheres*, **119,** 23-38.

Cai, X., Z.-L. Yang, J. B. Fisher, X. Zhang, M. Barlage, and F. Chen, 2016: Integration of nitrogen dynamics into the Noah-MP land surface model v1. 1 for climate and environmental predictions. *Geoscientific Model Development*, **9,** 1.

Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for EI Nino forecasting: implications for predictability. *Science*, **269,** 1699-1702.

Cho, Y., and B. A. Engel, 2017: NEXRAD quantitative precipitation estimations for hydrologic simulation using a hybrid hydrologic model. *Journal of Hydrometeorology*, **18,** 25-47.

Chou, C., 2003: Land–sea heating contrast in an idealized Asian summer monsoon. *Climate Dynamics*, **21,** 11-25.

Clark, M. P., and Coauthors, 2015: Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, **51,** 5929-5956.

Cocks, S. B., S. M. Martinaitis, B. Kaney, J. Zhang, and K. Howard, 2016: MRMS QPE performance during the 2013/14 cool season. *Journal of Hydrometeorology*, **17,** 791-810.

Corfidi, S., J. Meritt, and J. Fritsch, 1996: Predicting the movement of mesoscale convective complexes. *Weather and Forecasting*, **11,** 41-46.

Cunha, L. K., P. V. Mandapaka, W. F. Krajewski, R. Mantilla, and A. A. Bradley, 2012: Impact of radar–rainfall error structure on estimated flood magnitude across scales: An

investigation based on a parsimonious distributed hydrological model. *Water Resources Research*, **48,** W15015.

Dai, A., H. Li, Y. Sun, L.-C. Hong, LinHo, C. Chou, and T. Zhou, 2013: The relative roles of upper and lower tropospheric thermal contrasts and tropical influences in driving Asian summer monsoons. *Journal of Geophysical Research: Atmospheres*, **118,** 7024-7045.

David, C. H., Z.-L. Yang, and S. Hong, 2013: Regional-scale river flow modeling using off-the-shelf runoff products, thousands of mapped rivers and hundreds of stream flow gauges. *Environmental modelling & software*, **42,** 116-132.

David, C. H., J. S. Famiglietti, Z. L. Yang, and V. Eijkhout, 2015: Enhanced fixed‐size parallel speedup with the Muskingum method using a trans–boundary approach and a large subbasins approximation. *Water Resources Research*, **51,** 7547-7571.

David, C. H., D. R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout, 2011: River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology*, **12,** 913-934.

Deardorff, J. W., 1978: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research: Oceans*, **83,** 1889-1903.

Dee, D. P., and Coauthors, 2011: The ERA–Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, **137,** 553-597.

DelSole, T., and J. Shukla, 2010: Model fidelity versus skill in seasonal forecasting. *Journal of Climate*, **23,** 4794-4806.

Dickinson, E., A. Henderson-Sellers, and J. Kennedy, 1993: Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to the NCAR community climate model.

Dickinson, R. E., 1995: Land processes in climate models. *Remote Sensing of Environment*, **51,** 27-38.

Ding, Y., and J. C. Chan, 2005: The East Asian summer monsoon: an overview. *Meteorology and Atmospheric Physics*, **89,** 117-142.

Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki, and N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, **87,** 1381-1397.

Doblas–Reyes, F. J., J. Garcia–Serrano, F. Lienert, A. P. Biescas, and L. R. Rodrigues, 2013: Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, **4,** 245-268.

Douville, H., 2010: Relative contribution of soil moisture and snow mass to seasonal climate predictability: a pilot study. *Climate Dynamics*, **34,** 797-818.

Downton, M. W., J. Z. B. Miller, and R. A. Pielke Jr, 2005: Reanalysis of US National Weather Service flood loss database. *Natural Hazards Review*, **6,** 13-22.

Du, J., H. Xie, Y. Hu, Y. Xu, and C.-Y. Xu, 2009: Development and testing of a new storm runoff routing approach based on time variant spatially distributed travel time method. *Journal of Hydrology*, **369,** 44-54.

Durand, M., E. J. Kim, and S. A. Margulis, 2009: Radiance assimilation shows promise for snowpack characterization. *Geophysical Research Letters*, **36**.

Durand, M., L.-L. Fu, D. P. Lettenmaier, D. E. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez, 2010: The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, **98,** 766-779.

Eagleson, P. S., 1986: The emergence of global–scale hydrology. *Water Resources Research*, **22,** 6S-14S.

Eicker, A., M. Schumacher, J. Kusche, P. Döll, and H. M. Schmied, 2014: Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surveys in Geophysics*, **35,** 1285-1309.

Emanuel, K., 2017: Assessing the present and future probability of Hurricane Harvey's rainfall. *Proceedings of the National Academy of Sciences***,** 201716222.

Follum, M. L., A. A. Tavakoly, J. D. Niemann, and A. D. Snow, 2017: AutoRAPID: a model for prompt streamflow estimation and flood inundation mapping over regional to continental extents. *JAWRA Journal of the American Water Resources Association*, **53,** 280-299.

Furl, C., H. Sharif, J. W. Zeitler, A. E. Hassan, and J. Joseph, 2018: Hydrometeorology of the catastrophic Blanco river flood in South Texas, May 2015. *Journal of Hydrology: Regional Studies*, **15,** 90-104.

Furl, C., H. O. Sharif, A. E. Hassan, N. Mazari, D. Burtch, and G. L. Mullendore, 2015: Hydrometeorological Analysis of Tropical Storm Hermine and Central Texas Flash Flooding, September 2010. *Journal of Hydrometeorology*, **16,** 2311-2327.

Gates, W. L., 1992: AMIP: The atmospheric model intercomparison project. *Bulletin of the American Meteorological Society*, **73,** 1962-1970.

Giorgi, F., and R. Avissar, 1997: Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling. *Reviews of Geophysics*, **35,** 413-437.

Gleason, C. J., and L. C. Smith, 2014: Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, **111,** 4788-4791.

Gochis, D., W. Yu, and D. Yates, 2015: The WRF-Hydro Model Technical Description and User's Guide, Version 3.0, NCAR Technical Document, 120 pp.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, **21,** 1111-1152.

Goodall, J. L., and Coauthors, 2013: Coupling climate and hydrological models: Interoperability through Web Services. *Environmental Modelling & Software*, **46,** 250-259.

Goodchild, M. F., 1992: Geographical information science. *International journal of geographical information systems*, **6,** 31-45.

Goswami, B., V. Krishnamurthy, and H. Annmalai, 1999: A broad–scale circulation index for the interannual variability of the Indian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, **125,** 611-633.

Grice, G., and R. Maddox, 1983: Synoptic characteristics of heavy rainfall events in south Texas. *Preprints Fifth Conference on Hydrometeorology. Boston: American Meteorological Society*, 149-155.

Guo, Z., P. A. Dirmeyer, T. DelSole, and R. D. Koster, 2012: Rebound in atmospheric predictability and the role of the land surface. *Journal of Climate*, **25,** 4744-4749.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, **377,** 80-91.

Henderson-Sellers, A., Z.-L. Yang, and R. E. Dickinson, 1993: The Project for Intercomparison of Land-surface Parameterization Schemes. *Bulletin of the American Meteorological Society*, **74,** 1335-1350.

Hirsch, A. L., J. Kala, A. J. Pitman, C. Carouge, J. P. Evans, V. Haverd, and D. Mocko, 2014: Impact of land surface initialization approach on subseasonal forecast skill: A regional analysis in the Southern Hemisphere. *Journal of Hydrometeorology*, **15,** 300-319.

Hurrell, J. W., J. J. Hack, D. Shea, J. M. Caron, and J. Rosinski, 2008: A new sea surface temperature and sea ice boundary dataset for the Community Atmosphere Model. *Journal of Climate*, **21,** 5145-5153.

Jeong, J.-H., H. W. Linderholm, S.-H. Woo, C. Folland, B.-M. Kim, S.-J. Kim, and D. Chen, 2013: Impacts of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *Journal of Climate*, **26,** 1956-1972.

Jiang, X., G. Y. Niu, and Z. L. Yang, 2009: Impacts of vegetation and groundwater dynamics on warm season precipitation over the Central United States. *Journal of Geophysical Research: Atmospheres*, **114**.

Kauffeldt, A., F. Wetterhall, F. Pappenberger, P. Salamon, and J. Thielen, 2016: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environmental Modelling & Software*, **75,** 68-76.

Keim, B. D., R. A. Muller, and G. W. Stone, 2007: Spatiotemporal patterns and return periods of tropical storm and hurricane strikes from Texas to Maine. *Journal of climate*, **20,** 3498-3509.

Kim, J. H., Z. W. Geem, and E. S. Kim, 2001: Parameter estimation of the nonlinear Muskingum model using harmony search. *JAWRA Journal of the American Water Resources Association*, **37,** 1131-1138.

Kitoh, A., H. Endo, K. Krishna Kumar, I. F. A. Cavalcanti, P. Goswami, and T. Zhou, 2013: Monsoons in a changing world: A regional perspective in a global context. *Journal of Geophysical Research: Atmospheres*, **118,** 3053-3065.

Kitzmiller, D., D. Miller, R. Fulton, and F. Ding, 2013: Radar and multisensor precipitation estimation techniques in National Weather Service hydrologic operations. *Journal of Hydrologic Engineering*, **18,** 133-142.

Koster, R., and G. Walker, 2015: Interactive vegetation phenology, soil moisture, and monthly temperature forecasts. *Journal of Hydrometeorology*, **16,** 1456-1465.

Koster, R., and Coauthors, 2011: The second phase of the global land–atmosphere coupling experiment: soil moisture contributions to subseasonal forecast skill. *Journal of Hydrometeorology*, **12,** 805-822.

Koster, R. D., M. J. Suarez, and M. Heiser, 2000a: Variance and predictability of precipitation at seasonal-to-interannual timescales. *Journal of hydrometeorology*, **1,** 26-46.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000b: A catchment‐based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres*, **105,** 24809-24822.

Koster, R. D., S. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle, 2010a: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience*, **3,** 613.

Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *Journal of Climate*, **22,** 4322-4335.

Koster, R. D., and Coauthors, 2004a: Realistic Initialization of Land Surface States: Impacts on Subseasonal Forecast Skill. *Journal of Hydrometeorology*, **5,** 1049-1063.

Koster, R. D., and Coauthors, 2004b: Regions of strong coupling between soil moisture and precipitation. *Science*, **305,** 1138-1140.

Koster, R. D., and Coauthors, 2010b: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi‐model experiment. *Geophysical Research Letters*, **37**.

Kumar, S. V., P. A. Dirmeyer, C. D. Peters-Lidard, R. Bindlish, and J. Bolten, 2018: Information theoretic evaluation of satellite soil moisture retrievals. *Remote Sensing of Environment*, **204,** 392-400.

Kumar, S. V., and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *Journal of Hydrometeorology*, **15,** 2446-2469.

Kumar, S. V., and Coauthors, 2016: Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology*, **17,** 1951-1972.

Kwon, Y., Z.-L. Yang, T. J. Hoar, and A. M. Toure, 2017: Improving the radiance assimilation performance in estimating snow water storage across snow and land-cover types in north america. *Journal of Hydrometeorology*, **18,** 651-668.

Kwon, Y., A. M. Toure, Z.-L. Yang, M. Rodell, and G. Picard, 2015: Error characterization of coupled land surface-radiative transfer models for snow microwave radiance assimilation. *IEEE Transactions on Geoscience and Remote Sensing*, **53,** 5247-5268.

Kwon, Y., Z.-L. Yang, L. Zhao, T. J. Hoar, A. M. Toure, and M. Rodell, 2016: Estimating snow water storage in North America using CLM4, DART, and snow radiance data assimilation. *Journal of Hydrometeorology*, **17,** 2853-2874.

Lahoz, W., B. Khattatov, and R. Ménard, 2010: Data assimilation and information. *Data Assimilation*, Springer, 3-12.

Lahoz, W. A., and P. Schneider, 2014: Data assimilation: making sense of Earth Observation. *Frontiers in Environmental Science*, **2,** 16.

Lehner, B., and G. Grill, 2013: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, **27,** 2171-2186.

Leopold, L. B., 1991: Lag times for small drainage basins. *Catena*, **18,** 157-171.

Lettenmaier, D. P., D. Alsdorf, J. Dozier, G. J. Huffman, M. Pan, and E. F. Wood, 2015: Inroads of remote sensing into hydrologic science during the WRR era. *Water Resources Research*, **51,** 7309-7342.

Li, H., M. S. Wigmosta, H. Wu, M. Huang, Y. Ke, A. M. Coleman, and L. R. Leung, 2013: A physically based runoff routing model for land surface and earth system models. *Journal of Hydrometeorology*, **14,** 808-828.

Li, Y., Y. Ding, and W. Li, 2017: Interdecadal variability of the Afro-Asian summer monsoon system. *Advances in Atmospheric Sciences*, **34,** 833-846.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, **99,** 14415-14428.

Lin, P., Z.-L. Yang, X. Cai, and C. H. David, 2015: Development and evaluation of a physically-based lake level model for water resource management: A case study for Lake Buchanan, Texas. *Journal of Hydrology: Regional Studies*, **4,** 661-674.

Lin, P., J. Wei, Z. L. Yang, Y. Zhang, and K. Zhang, 2016: Snow data assimilation‐constrained land initialization improves seasonal temperature prediction. *Geophysical Research Letters*, **43**.

Lin, P., and Coauthors, 2018: Spatiotemporal Evaluation of Simulated Evapotranspiration and Streamflow over Texas Using the WRF–Hydro–RAPID Modeling Framework. *JAWRA Journal of the American Water Resources Association*, **54,** 40-54.

Lin, Y., 2011: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data, Version 1.0, UCAR/NCAR–Earth Observing Laboratory.

Liu, F., and B. R. Hodges, 2014: Applying microprocessor analysis methods to river network modelling. *Environmental Modelling & Software*, **52,** 234-252.

Liu, Y., C. D. Peters–Lidard, S. V. Kumar, K. R. Arsenault, and D. M. Mocko, 2015: Blending satellite‐based snow depth products with in situ observations for streamflow predictions in the Upper Colorado River Basin. *Water resources research*, **51,** 1182-1202.

Liu, Y. Y., D. R. Maidment, D. G. Tarboton, X. Zheng, and S. Wang, 2017: A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping. *Journal of the American Water Resources Association (JAWRA)*.

Lohmann, D., E. Raschke, B. Nijssen, and D. Lettenmaier, 1998: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrological Sciences Journal*, **43,** 131-141.

Looper, J. P., and B. E. Vieux, 2012: An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model. *Journal of Hydrology*, **412,** 114-132.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on predictability*.

Lu, M., S. Yang, Z. Li, B. He, S. He, and Z. Wang, 2017: Possible effect of the Tibetan Plateau on the “upstream” climate over West Asia, North Africa, South Europe and the North Atlantic. *Climate Dynamics***,** 1-14.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and meso-α scale aspects of flash flood events. *Bulletin of the American Meteorological Society***,** 115-123.

Mahanama, S., B. Livneh, R. Koster, D. Lettenmaier, and R. Reichle, 2012: Soil moisture, snow, and seasonal streamflow forecasts in the United States. *Journal of Hydrometeorology*, **13,** 189-203.

Maidment, D. R., 2017: Conceptual framework for the national flood interoperability experiment. *JAWRA Journal of the American Water Resources Association*, **53,** 245-257.

Manabe, S., 1969: CLIMATE AND THE OCEAN CIRCULATION. *Monthly Weather Review*, **97,** 739-774.

Maxwell, R. M., and N. L. Miller, 2005: Development of a Coupled Land Surface and Groundwater Model. *Journal of Hydrometeorology*, **6,** 233-247.

McDonnell, J. J., and K. Beven, 2014: Debates–The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, **50,** 5342-5350.

Merkel, W. H., 2002: Muskingum-Cunge flood routing procedure in NRCS hydrologic models. *Second Fed. Interag. Hydrol. Model. Conf., Las Vegas, Nevada. July*.

Merwade, V., F. Olivera, M. Arabi, and S. Edleman, 2008: Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions. *Journal of Hydrologic Engineering*, **13,** 608-620.

Milly, P. C., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer, 2008: Stationarity is dead: Whither water management? *Science*, **319,** 573-574.

Mizukami, N., and Coauthors, 2016: mizuRoute version 1: a river network routing tool for a continental domain water resources applications. *Geoscientific Model Development*, **9,** 2223.

Moore, R. B., and T. G. Dewald, 2016: The road to NHDPlus–advancements in digital stream networks and associated catchments. *JAWRA Journal of the American Water Resources Association*, **52,** 890-900.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams, 2011: Soil and water assessment tool theoretical documentation version 2009.

Niu, G. Y., and Z. L. Yang, 2007: An observation‐based formulation of snow cover fraction and its evaluation over large North American river basins. *Journal of Geophysical Research: Atmospheres*, **112**.

Niu, G. Y., Z. L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su, 2007: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research: Atmospheres*, **112**.

Niu, G. Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah–MP): 1. Model description and evaluation with local‐scale measurements. *Journal of Geophysical Research: Atmospheres*, **116**.

Oleson, K. W., and Coauthors, 2010: Technical description of version 4.0 of the Community Land Model (CLM).

Orsolini, Y., and Coauthors, 2013: Impact of snow initialization on sub-seasonal forecasts. *Climate dynamics*, **41,** 1969-1982.

Palmer, T. N., and D. L. Anderson, 1994: The prospects for seasonal forecasting—A review paper. *Quarterly Journal of the Royal Meteorological Society*, **120,** 755-793.

Pielke, R. A., M. W. Downton, and J. B. Miller, 2002: *Flood damage in the United States, 1926-2000: a reanalysis of National Weather Service estimates.* University Corporation for Atmospheric Research Boulder, CO.

Pitman, A. J., 2003: The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, **23,** 479-510.

Qi, Y., S. Martinaitis, J. Zhang, and S. Cocks, 2016: A Real–Time Automated Quality Control of Hourly Rain Gauge Data Based on Multiple Sensors in MRMS System. *Journal of Hydrometeorology*, **17,** 1675-1691.

Rajib, M. A., V. Merwade, and Z. Yu, 2016: Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*, **536,** 192-207.

Reges, H. W., N. Doesken, J. Turner, N. Newman, A. Bergantino, and Z. Schwalbe, 2016: COCORAHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society*, **97,** 1831-1846.

Rodell, M., and Coauthors, 2004: The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85,** 381-394.

Saha, S., and Coauthors, 2014: The NCEP climate forecast system version 2. *Journal of Climate*, **27,** 2185-2208.

Saharia, M., P.-E. Kirstetter, H. Vergara, J. J. Gourley, Y. Hong, and M. Giroud, 2017: Mapping flash flood severity in the United States. *Journal of Hydrometeorology*, **18,** 397-411.

Salas, F. R., and Coauthors, 2018: Towards Real–Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. *JAWRA Journal of the American Water Resources Association*, **54,** 7-27.

Santanello Jr, J. A., S. V. Kumar, C. D. Peters-Lidard, and P. M. Lawston, 2016: Impact of soil moisture assimilation on land surface model spinup and coupled land–atmosphere prediction. *Journal of Hydrometeorology*, **17,** 517-540.

Scanlon, B. R., and Coauthors, 2018: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*.

Schaake, J. C., V. I. Koren, Q. Y. Duan, K. Mitchell, and F. Chen, 1996: Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of Geophysical Research: Atmospheres*, **101,** 7461-7475.

Schroeder, A., J. Basara, J. M. Shepherd, and S. Nelson, 2016: Insights into atmospheric contributors to urban flash flooding across the United States using an analysis of rawinsonde data and associated calculated parameters. *Journal of Applied Meteorology and Climatology*, **55,** 313-323.

Schumacher, R. S., and R. H. Johnson, 2005: Organization and environmental properties of extreme-rain-producing mesoscale convective systems. *Monthly weather review*, **133,** 961-976.

——, 2009: Quasi-stationary, extreme-rain-producing convective systems associated with midlevel cyclonic circulations. *Weather and Forecasting*, **24,** 555-574.

Sellers, P., Y. Mintz, Y. e. a. Sud, and A. Dalcher, 1986: A simple biosphere model (SiB) for use within general circulation models. *Journal of the Atmospheric Sciences*, **43,** 505-531.

Senan, R., and Coauthors, 2016: Impact of springtime Himalayan–Tibetan Plateau snowpack on the onset of the Indian summer monsoon in coupled seasonal forecasts. *Climate Dynamics*, **47,** 2709-2725.

Senatore, A., G. Mendicino, D. J. Gochis, W. Yu, D. N. Yates, and H. Kunstmann, 2015: Fully coupled atmosphere‐hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *Journal of Advances in Modeling Earth Systems*, **7,** 1693-1715.

Seol, K.-H., and S.-Y. Hong, 2009: Relationship between the Tibetan Snow in Spring and the East Asian Summer Monsoon in 2003: A Global and Regional Modeling Study. *Journal of Climate*, **22,** 2095-2110.

Sharif, H. O., A. A. Hassan, S. Bin-Shafique, H. Xie, and J. Zeitler, 2010: Hydrologic Modeling of an Extreme Flood in the Guadalupe River in Texas1. *JAWRA Journal of the American Water Resources Association*, **46,** 881-891.

Shukla, J., 1998: Predictability in the midst of chaos: A scientific basis for climate forecasting. *science*, **282,** 728-731.

Shukla, J., and Coauthors, 2000: Dynamical seasonal prediction. *Bulletin of the American Meteorological Society*, **81,** 2593-2606.

Silver, M., A. Karnieli, H. Ginat, E. Meiri, and E. Fredj, 2017: An innovative method for determining hydrological calibration parameters for the WRF-Hydro model in arid regions. *Environmental Modelling & Software*, **91,** 47-69.

Singh, R., J. Reager, N. Miller, and J. Famiglietti, 2015: Toward hyper–resolution land–surface modeling: The effects of fine–scale topography and soil texture on CLM4. 0 simulations over the Southwestern US. *Water Resources Research*, **51,** 2648-2667.

Singh, V. P., and L. Zhang, 2008: At–a–station hydraulic geometry relations, 1: theoretical development. *Hydrological Processes*, **22,** 189-215.

Sivapalan, M., 2018: From engineering hydrology to Earth system science: milestones in the transformation of hydrologic science. *Hydrology and Earth System Sciences*, **22,** 1665.

Slater, A. G., and M. P. Clark, 2006: Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, **7,** 478-493.

Smith, J. A., M. L. Baeck, J. E. Morrison, and P. Sturdevant-Rees, 2000: Catastrophic rainfall and flooding in Texas. *Journal of Hydrometeorology*, **1,** 5-25.

Smith, M. B., and Coauthors, 2012: Results of the DMIP 2 Oklahoma experiments. *Journal of hydrology*, **418,** 17-48.

Su, F., X. Duan, D. Chen, Z. Hao, and L. Cuo, 2013: Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau. *Journal of Climate*, **26,** 3187-3208.

Su, H., Z. L. Yang, R. E. Dickinson, C. R. Wilson, and G. Y. Niu, 2010: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information. *Journal of Geophysical Research: Atmospheres*, **115**.

Sun, C., J. P. Walker, and P. R. Houser, 2004: A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research: Atmospheres*, **109**.

Sun, Y., L. Gu, R. E. Dickinson, R. J. Norby, S. G. Pallardy, and F. M. Hoffman, 2014: Impact of mesophyll diffusion on estimated global land $CO_2$ fertilization. *Proceedings of the National Academy of Sciences*, **111,** 15774-15779.

Tavakoly, A. A., A. D. Snow, C. H. David, M. L. Follum, D. R. Maidment, and Z. L. Yang, 2017: Continental–Scale River Flow Modeling of the Mississippi River Basin Using High–Resolution NHDPlus Dataset. *JAWRA Journal of the American Water Resources Association*, **53,** 258-279.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106,** 7183-7192.

Thomas, J. A., A. A. Berg, and W. J. Merryfield, 2016: Influence of snow and soil moisture initialization on sub-seasonal predictability and forecast skill in boreal spring. *Climate dynamics*, **47,** 49-65.

Todini, E., 2007: A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. *Hydrology and Earth System Sciences*, **11,** 1645-1659.

Toure, A. M., M. Rodell, Z.-L. Yang, H. Beaudoing, E. Kim, Y. Zhang, and Y. Kwon, 2016: Evaluation of the snow simulations from the Community Land Model, version 4 (CLM4). *Journal of Hydrometeorology*, **17,** 153-170.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bulletin of the American Meteorological Society*, **84,** 1205-1217.

Turner, A. G., and J. M. Slingo, 2011: Using idealized snow forcing to test teleconnections with the Indian summer monsoon in the Hadley Centre GCM. *Climate Dynamics*, **36,** 1717-1735.

Veilleux, A. G., T. A. Cohn, K. M. Flynn, R. R. Mason Jr, and P. R. Hummel, 2014: Estimating magnitude and frequency of floods using the PeakFQ 7.0 program2327-6932.

Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quarterly Journal of the Royal Meteorological Society*, **118,** 877-926.

Webster, P. J., and C. D. Hoyos, 2010: Beyond the spring barrier? *Nature Geoscience*, **3,** 152.

Webster, P. J., V. O. Magana, T. Palmer, J. Shukla, R. Tomas, M. Yanai, and T. Yasunari, 1998: Monsoons: Processes, predictability, and the prospects for prediction. *Journal of Geophysical Research: Oceans*, **103,** 14451-14510.

Williams, I. N., and M. S. Torn, 2015: Vegetation controls on surface heat flux partitioning, and land‐atmosphere coupling. *Geophysical Research Letters*, **42,** 9416-9424.

Wood, E. F., 1991: Global Scale Hydrology: Advances in Land Surface Modeling. *Reviews of Geophysics*, **29,** 193-201.

Wood, E. F., and Coauthors, 2011: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, **47**.

Wright, D. B., J. A. Smith, and M. L. Baeck, 2014: Flood frequency analysis using radar rainfall fields and stochastic storm transposition. *Water Resources Research*, **50,** 1592-1615.

Wu, G., Y. Liu, B. He, Q. Bao, A. Duan, and F.-F. Jin, 2012: Thermal controls on the Asian summer monsoon. *Scientific reports*, **2,** 404.

Wu, G., and Coauthors, 2007: The influence of mechanical and thermal forcing by the Tibetan Plateau on Asian climate. *Journal of Hydrometeorology*, **8,** 770-789.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, **117,** n/a-n/a.

Xu, L., and P. Dirmeyer, 2011: Snow–atmosphere coupling strength in a global atmospheric model. *Geophysical Research Letters*, **38**.

——, 2013: Snow–atmosphere coupling strength. Part II: Albedo effect versus hydrological effect. *Journal of Hydrometeorology*, **14,** 404-418.

Xue, Y., F. De Sales, R. Vasic, C. R. Mechoso, A. Arakawa, and S. Prince, 2010: Global and seasonal assessment of interactions between climate and vegetation biophysical processes: A GCM study with different land–vegetation representations. *Journal of Climate*, **23,** 1411-1433.

Yamazaki, D., G. A. Almeida, and P. D. Bates, 2013: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector‐based river network map. *Water Resources Research*, **49,** 7221-7235.

Yang, Z.-L., 2004: Modeling land surface processes in short-term weather and climate studies. *Observation, Theory and Modeling of Atmospheric Variability*, WORLD SCIENTIFIC, 288-313.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research: Atmospheres*, **116,** n/a-n/a.

Yasunari, T., A. Kitoh, and T. Tokioka, 1991: Local and remote responses to excessive snow mass over Eurasia appearing in the northern spring and summer climate. *Journal of the Meteorological Society of Japan. Ser. II*, **69,** 473-487.

Yilmaz, K. K., H. V. Gupta, and T. Wagener, 2008: A process‐based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, **44**.

Yucel, I., A. Onen, K. Yilmaz, and D. Gochis, 2015: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *Journal of Hydrology*, **523,** 49-66.

Zhang, J., Y. F. Huang, D. Munasinghe, Z. Fang, Y. P. Tsang, and S. Cohen, 2018: Comparative Analysis of Inundation Mapping Approaches for the 2016 Flood in the Brazos River, Texas. *JAWRA Journal of the American Water Resources Association*, **0**.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. *Bulletin of the American Meteorological Society*, **97,** 621-638.

Zhang, Y., 2015: Multivariate land snow data assimilation in the Northern Hemisphere: Development, evaluation and uncertainty quantification of the extensible data assimilation system.

Zhang, Y. F., and Z. L. Yang, 2016: Estimating uncertainties in the newly developed multi‐source land snow data assimilation system. *Journal of Geophysical Research: Atmospheres*, **121,** 8254-8268.

Zhang, Y. F., T. J. Hoar, Z. L. Yang, J. L. Anderson, A. M. Toure, and M. Rodell, 2014: Assimilation of MODIS snow cover through the Data Assimilation Research Testbed and the Community Land Model version 4. *Journal of Geophysical Research: Atmospheres*, **119,** 7091-7103.

Zhang, Z., V. Koren, S. Reed, M. Smith, Y. Zhang, F. Moreda, and B. Cosgrove, 2012: SAC-SMA a priori parameter differences and their impact on distributed hydrologic model simulations. *Journal of hydrology*, **420,** 216-227.

Zhao, K., and M. Xue, 2009: Assimilation of coastal Doppler radar data with the ARPS 3DVAR and cloud analysis for the prediction of Hurricane Ike (2008). *Geophysical Research Letters*, **36**.

Zhao, L., Z.-L. Yang, and T. J. Hoar, 2016: Global soil moisture estimation by assimilating AMSR-E brightness temperatures in a coupled CLM4–RTM–DART system. *Journal of Hydrometeorology*, **17,** 2431-2454.

Zhao, L., and Z-L. Yang, 2018 (*in revision*): Multi-Sensor Land Data Assimilation: Toward a Robust Global Soil Moisture and Snow Estimation, *Remote Sensing of Environment.*

Zheng, H., and Z. L. Yang, 2016: Effects of soil‐type datasets on regional terrestrial water cycle simulations under different climatic regimes. *Journal of Geophysical Research: Atmospheres*, **121**.

Zheng X., D.G. Tarboton, D. R. Maidment, Y.Y. Liu, and P. Passalacqua, 2018 (*in revision*): River channel geometry and rating curve estimation using height above the nearest drainage, *J. Am. Water. Resour. Assoc.*

Zhou, T., and L. Zou, 2010: Understanding the predictability of East Asian summer monsoon from the reproduction of land–sea thermal contrast change in AMIP-type simulation. *Journal of Climate*, **23,** 6009-6026.

# Vita

Peirong Lin was born in Yunnan Province, the beautiful southwestern China in 1990. After graduating from the Middle School attached to Yunnan Normal University in 2008, she entered the School of Earth and Space Science (SESS) in Peking University, Beijing, for her undergraduate training. She studied general earth sciences during the freshman and sophomore years, before she focused the major field of study in Geographic Information Science (GIS) and Remote Sensing (RS). She earned the B.S. degree in the year of 2012, after which she joined the University of Texas at Austin to pursue a Ph.D. degree in Climate System Science. She will continue as a geoscientist in academia.

Permanent email address: peironglinlin@gmail.com.

This dissertation was typed by Peirong Lin.